105
A-n 13

8 No. 1 (New Series No. 217) January 1988

ANALYSIS

Cole-A 02958-17-P004462
17

Edited by

PETER SMITH

CONTENTS

rial	
's first empiricist criterion of meaning: Why does it fail?	DAVID LEWIS
sistency enough for existence in mathematics?	GEOFFREY HUNTER
' and two types of quantification	D. H. RICE
argument of Armstrong against Nominalism	MICHAEL LEVIN
s's modal realism: a reply to Naylor	MARK F. SHARLOW
te on objects and events	OLAV GJELSVIK
ragmatic problem of induction	JOHN WATKINS
of sphexishness	RICHARD DOUBLE
on the biological basis of cognition	ASHER SEIDEL
g on Parfit's relation R	MARK SIDERITS
is there something rather than nothing?	NOEL FLEMING
cability and the unpreventable	N. M. L. NATHAN
Mothersill on aesthetic pleasure	WILLIAM CHARLTON
tion, foresight and collective responsibility	GREGORY MELLEMA
on tracing mental images	DAVID GORDON
ng Superman again: a reply to Clark	SELMER BRINGSJORD
rman and the duck/rabbit: a reply to Gordon and Bringsjord	ANDY CLARK
an on evidential insatiability	HOWARD SIMMONS
an on global justification	ROBERT P. McARTHUR
ms on what the president knew	THOMAS P. FLINT
much did the president know?	C. J. F. WILLIAMS

BASIL BLACKWELL · OXFORD AND NEW YORK

ISSN 0003-2638

BLACKWELL

Colour

Some Philosophical Problems from Wittgenstein

JONATHAN WESTPHAL

The author formulates an important new theory of colour and its perception: if he is correct, the way we understand what it is to talk about colour will confound most attempts to characterize the divide between mental and physical.

Aristotelian Society Series
128 pages, **£19.50** (0 631 14937 6)

Aesthetic Reconstructions

The Seminal Writings of Lessing, Kant and Schiller

ANTHONY SAVILE

'A highly intelligent and exceptionally well-argued work that will provide a fresh impetus to the serious study of the aesthetic problems that occupied Lessing, Kant and Schiller.' Flint Schier, *University of Glasgow*

Aristotelian Society Series
272 pages, **£27.50** (0 631 158719 7)

The Logic of Mortality

ANTONY FLEW

Examining and questioning the assumptions on which immortality theories are based, Professor Flew presents a vigorous and controversial analysis of our mortality and identity.
216 pages, **£22.50** (0 631 15628 3)

Continental Philosophy and Modern Theology

An Engagement

DAVID BROWN

In the first systematic survey of this area, David Brown shows how the major debates in theology have been influenced by continental philosophy and how the introduction of key thinkers to these debates can effect considerable reconciliations.
264 pages, **£27.50** (0 631 15734 4)

Science Without Unity

Reconciling the Human and Natural Sciences

JOSEPH MARGOLIS

Professor Margolis analyses in detail the conceptual relationships between the human and natural sciences, offering explicit, systematically related alternatives to the various expressions and developments of the unity model.
492 pages, **£29.50** (0 631 15173 7)

Basil Blackwell
108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016.
Toll-free ordering: 1-800-638-3030 (USA)

P 4462

105
Apr 13

PHILOSOPHICAL TOPICS

Editor: Christopher S. Hill

Volume 15, No. 1 Spring 1987
Topic: **Philosophy of Language**

Ernest W. Adams, "On the Meaning of the Conditional"; **Anthony Appiah,** "Why Componentiality Fails"; **Akeel Bilgrami,** "An Externalist Account of Psychological Content"; **Robert B. Brandom,** "Singular Terms and Sentential Sign Designs"; **Brian Loar,** "Subjective Intentionality"; **William G. Lycan,** "The Myth of the 'Projection Problem for Presuppostion' "; **Scott Soames,** "Direct Reference, Propositional Attitudes, and Semantic Content"; **Robert Stalnaker,** "Semantics for Belief"; **Crispin Wright,** "Further Reflections on the Sorites Paradox"

Volume 15, No. 2 Fall 1987
Topic: **Ancient Greek Philosophy**
Issue Editor: **Lynne Spellman**

Julia Annas, "Epicurus on Pleasure and Happiness"; **Scott Austin,** "How Did Socrates Know That He Didn't Know?"; **Sarah Waterlow Broadie,** "Nature, Craft and Phronesis in Aristotle"; **Alan Code,** "Soul as Efficient Cause in Aristotle's Embryology"; **Russell Dancy,** "Theaetetus' First Baby: Plato, *Theaetetus* 151e-160e"; **Terence Irwin,** "Ways to First Principles: Aristotle's Methods of Inquiry"; **Alexander Mourelatos,** "Gorgias on the Function of Language"; **Susan Sauve,** "Self-Movement and Moral Responsibility"; **Nicholas White,** "Plato and Frege: Some Comparisons"

Philosophical Topics appears twice yearly. Each issue is devoted to questions in one of the major areas of philosophy and consists entirely of solicited papers.

Subscription Rates: Domestic Individual Rate (U.S. and Canada)–$20; Foreign Individual Rate–$25 in U.S. Currency; Domestic Institutional Rate (U.S. and Canada)–$30; Foreign Institutional Rate–$35 in U.S. Currency. All checks should be made payable to The University of Arkansas Foundation, Inc./ *Philosophical Topics* and sent to the Managing Editor at the Department of Philosophy, 127 Kimpel Hall, University of Arkansas, Fayetteville, Arkansas 72701.

AUSTRALASIAN JOURNAL OF PHILOSOPHY
Contents of Vol. 65, No. 4; December 1987
EDITOR: BRIAN ELLIS

ARTICLES

Mark Leon Character, Content, and the Ontology of Experience
Janet Levin Physicalism and the Subjectivity of Secondary Qualities
Martin M. Tweedale Aristotle's Universals
Russell Wahl Some Consequences of Possibilism
Mark Heller The Best Candidate Approach to Diachronic Identity
John Bishop Sensitive and Insensitive Responses to Deviant Action
Roderic A. Girle The Concept of Revelation

CRITICAL NOTICE

Paul Simpson The Disappearance of Introspection

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTION: Individuals A$20 (US$20 or Stg 10); Institutions A$40, (US$40 or Stg 20). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

The Monist *An International Quarterly Journal of General Philosophical Inquiry*

Editor, John HOSPERS

Editorial Board: Henry Allison, William P. Alston, John M. Cooper, Dagfinn Føllesdal, R. Barcan Marcus, Joseph Margolis, Wallace I. Matson, Joseph Owens, Bruce Vermazen, Wilfrid Sellars, Anita Silvers, John E. Smith, Richard Wasserstrom.

Managing Editor: Sherwood J. B. SUGDEN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate in advance with the Editor for special instructions. Two nonreturnable copies (one copy outside the U.S. and Canada) should be submitted NINE MONTHS PRIOR to the announced publiation date. Papers should be prepared for anonymous submission to readers, and should be from 10 to 20 double-spaced typewritten pages in length, including notes.

SCHEDULED PUBLICATION DATES:

Vol. 71, No. 1	Jan.,	1988	Philosophical Problems of Space Exploration
Vol. 71, No. 2	Apr.,	1988	Aesthetics and the Histories of the Arts
Vol. 71, No. 3	July,	1988	Justification
Vol. 71, No. 4	Oct.,	1988	Descartes and His Contemporaries
Vol. 72, No. 1	Jan.,	1989	Logic as a Field of Knowledge
Vol. 72, No. 2	Apr.,	1989	Kant's Critical Philosophy
Vol. 72, No. 3	July,	1989	Kant's Practical Philosophy
Vol. 72, No. 4	Oct.,	1989	The Philosophy of R. G. Collingwood
Vol. 73, No. 1	Jan.,	1990	Hellenistic Ethics

Editorial Office: School of Philosophy, University of Southern California
Los Angeles, CA 90089
Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, $24.00; Individuals, $15.00. Single copies: $6.00. Foreign postage: Add 50 cents to single copy rate or $1.00 to subscription.

P 4462

INQUIRY
An Interdisciplinary Journal of Philosophy

Editor: Alastair Hannay

Volume 30 (1987), Nos. 4

How to Believe in Fairies	Stephen R. L. Clark
The Origins of Folk Psychology	George Graham
The Very Idea of a Folk Psychology	R. A. Sharpe
The Claims of Consciousness: A Critical Survey	Alastair Hannay
Moral Doubts about Strict Materialism	Torbjörn Tännsjö
A Kripkean Objection to Kripke's Argument against Identity-theories	Olav Gjelsvik
Coming of Age in the Philosophy of Biology (Elliott Sober: *The Nature of Selection: Evolutionary Theory in Philosophical Focus*)	Michael Bradie

---

Subscription price:
Institutions: USD 60.-. Individuals: USD 31.-.

Published quarterly by Norwegian University Press, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway - Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003, USA. Editorial address: Institute of Philosophy, P.O. Box 1024 Blindern, 0315 Oslo 3, Norway.

ANALYSIS 48.1 JANUARY 1988

GRANTS IN SUPPORT OF PHILOSOPHY CONFERENCES

THE ANALYSIS committee has agreed to make limited funds available to support philosophy conferences held in the United Kingdom. These funds are to help pay for the travelling expenses of invited speakers whose expenses cannot be funded from other sources. (Such speakers will typically be people invited from elsewhere in the U.K., as the British Academy may offer support for the expenses of people invited from abroad.) The sum available will normally be no more than £200 per conference.

In order to gain support, a conference must be organized in such a way that attendance is not narrowly restricted, and the following conditions must be met:

- (i) attendance at the conference should not be by invitation only, nor be limited to members of one institution,
- (ii) the conference should either take place out of normal university term time, or should take place at a weekend,
- (iii) the conference should be advertised widely (e.g. at least to a number of U.K. philosophy departments).

Applications for grants will be considered at the start of July and at the end of the year. Applications should be made to the Secretary of the ANALYSIS committee, Dr T. R. Baldwin, Clare College, Cambridge CB2 1TL.

AYER'S FIRST EMPIRICIST CRITERION OF MEANING: WHY DOES IT FAIL?

By DAVID LEWIS

IN the first edition of *Language, Truth and Logic*, Ayer proposes that 'the mark of a genuine factual proposition' is 'that some experiential propositions can be deduced from it in conjunction with certain other premises without being deducible from those other premises alone' ([1], pp. 38–9).

Berlin objects that the criterion is 'a good deal too liberal' and admits patent nonsense. His example, near enough, is that the experiential proposition 'I dislike this logical problem' can be deduced from the nonsensical 'this logical problem is green' in conjunction with the premise 'I dislike whatever is green' without being deducible from the latter premise alone ([2], p. 234). Berlin's point is well taken, if indeed a category mistake is nonsense. But it shows at most that the criterion admits too much. We do not yet know how much too much.

In his introduction to the second edition, Ayer accepts Berlin's point in a greatly extended form. Not only does the criterion admit this or that piece of nonsense; it 'allows meaning to any statement whatsoever. For, given any statement "*S*" and an observation-statement "*O*", "*O*" follows from "*S*" and "if *S* then *O*" without following from "if *S* then *O*" alone' ([1], p. 11).

Here Ayer goes wrong. For it may very well happen that the consequent of a conditional 'if *S* then *O*' does follow from that conditional alone. For instance, *O* follows from 'if (*P* or not *P*) then *O*' and from 'if not *O*, then *O*'. Ayer has just overlooked such cases. In general, *O* follows from 'if *S* then *O*' just when '*S* or *O*' is analytic.

(Two remarks. (1) I assume that Ayer's 'if... then' is meant truth-functionally. (2) I assume that 'follows' and 'analytic' are to be taken in corresponding senses, so that a conclusion follows from a premise just when the conditional from premise to conclusion is analytic. Then if 'analytic' covers more than narrowly logical truth, 'follows' must likewise cover more than narrowly logical implication. Alternatively, 'follows' can be glossed as 'follows by logic alone' provided that 'analytic' is glossed as 'logically true'. My point goes through either way.)

It should have been easy to see that the criterion does not admit just any statement as factual. For if *S* is analytic, nothing at all follows from *S* in conjunction with other premises, except for what follows from those other premises alone. To be sure, Ayer does count analytic statements as meaningful. But they are definitely not supposed to come out as factual; rather, the meaningful divides into the analytic and the factual ([1], p. 31).

Ayer is right, however, that the criterion 'allows meaning to any statement whatsoever'. Only his proof is at fault, not its conclusion. Any statement *S* is either factual by the criterion or else analytic, and counts as meaningful either way. The correct proof is by cases. First case: for some observation-statement *O*, '*S* or *O*' is not analytic. Then *O* does not follow from 'if *S* then *O*' alone, but does follow from *S* in conjunction with 'if *S* then *O*'. Then *S* is factual by the criterion. Second case: there is no such *O*. It is safe to assume that some two observation-statements O_1 and O_2 conflict, so that 'not both O_1 and O_2' is analytic. Also '*S* or O_1' and '*S* or O_2' are analytic *ex hypothesi*, and from these three analytic premises *S* follows. Then *S* itself is analytic. Q.E.D. Ayer's mistake was simply to omit the second case.

Later authors have perpetuated the mistake. Church writes that in Ayer's second edition, 'A criticism by Berlin ... is accepted as valid, according to which the definition as actually given in the first edition would make all statements verifiable' ([3], p. 52). Church himself implicitly endorses the criticism thus stated; yet it is mistaken, since Ayer uses 'verifiable' interchangeably with 'factual'. Further, Berlin said nothing so general. Several others paraphrase and endorse Ayer's faulty argument and conclusion: Hempel ([5], p. 49), Scheffler ([6], p. 41), and recently Foster ([4], p. 14).

Princeton University,
Princeton, NJ 08544, U.S.A.

© DAVID LEWIS 1988

REFERENCES

[1] A. J. Ayer, *Language, Truth and Logic*, Gollancz, 1936; second edition, 1946. Page references are to the second edition.

[2] Isaiah Berlin, 'Verification', *Proceedings of the Aristotelian Society*, 39 (1938-39).

[3] Alonzo Church, review of the second edition of *Language, Truth and Logic, Journal of Symbolic Logic*, 14 (1949).

[4] John Foster, *Ayer*, Routledge and Kegan Paul, 1985.

[5] Carl G. Hempel, 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie*, 4 (1950).

[6] Israel Scheffler, *The Anatomy of Inquiry*, Knopf, 1963.

IS CONSISTENCY ENOUGH FOR EXISTENCE IN MATHEMATICS?

By GEOFFREY HUNTER

HILBERT and Poincaré at various times said so:

> Hilbert: 'If the arbitrarily given axioms do not contradict one another, then they are true and the things defined by them exist. This for me is the criterion of truth and existence.' (Letter to Frege, 1899)
>
> Poincaré: 'What is the meaning in mathematics of the phrase *to exist*? It means, I said, to be free from contradiction.' (*Science and Method*, 1908, Bk 2, ch. V, pt. III)

Antony Flew asserted it in a discussion of the ontological argument in 1966: '... [Proofs of] the so-called "existence theorems" of [pure] mathematics ... are not proofs of the actual existence of anything, but only and properly of the freedom from contradiction of concepts.' (*God and Philosophy*, p. 79)

Outside mathematics we would distinguish *being free from contradiction* from *applying to something*. Why not inside? Hilbert, Poincaré and Flew don't give any reason. (John Derrick has suggested to me that for Hilbert at that time and for Poincaré the standard way of proving consistency was by producing a model. That would account for their equation of the two notions of *being consistent* and *having a model*.)

But the following proof of an existential theorem of mathematics is not (simply) a proof that a concept is free from contradiction (though it does prove that); rather, it is a proof that the concept applies to something, viz. the number 2.

Theorem: There exists an even prime number.
Proof: 2 is even.
2 is a prime number.
So there exists an even prime number.

It might be thought that the position since 1960 with respect to infinitesimals lends some support to the thesis that in pure mathematics consistency is enough for existence. Let us look at the situation more closely.

Take any set of first-order axioms adequate for the algebra of the real numbers. Call them R. To R add a new constant 'c' and denumerably many new axioms as follows:

$(\mathrm{E}x)\ x = c$
$c > 0$
$c < 1/2$
$c < 1/3$
$c < 1/4$
$c < 1/5$
....
....

Let the resulting system be R*. R* is a first-order theory.

For any finite subset of the new set of axioms you can find a real number which has the properties ascribed to c by that finite set of axioms, and which also has all the properties ascribed by the axioms of R, understood in their ordinary intended sense.

So: every finite subset of the axioms of R* has a model.

So by the Compactness Theorem (*If every finite subset of the proper axioms of a first-order theory K has a model, then K has a model*) R* has a model.

Since R* has a model, R* is formally consistent.

Any model of R* will have as a member of its domain something that

(i) obeys formal analogues of every (elementarily expressible) law that every real number obeys, but
(ii) is also *greater than whatever in the model corresponds to '0' and *less than everything in the model that is formally

> analogous to an ordinary positive real number (where **greater than* and **less than* are formal analogues of ordinary *greater than* and *less than* respectively).

Call such a thing an *infinitesimal. There will be in the model infinitely many *infinitesimals, since it will be true of whatever in the model is denoted by '*c*' that there is something *greater than *0 and *less than *c*, and something else again *less than this new entity yet *greater than *0, and so on, since it is true of any real number greater than 0 that there is a real number greater than 0 but less than it.

(It may be of help to the reader to say here that the axioms of R* ensure that *c* is *less than every *positive rational, and for every positive real there is a positive rational less than it, so *c* will be *less than every *positive real.)

So here we have a consistent theory of things that behave just the way some seventeenth, eighteenth and nineteenth century mathematicians wanted their infinitely small quantities to behave. Among the propositions of the consistent theory is one asserting the existence of such a thing (I put it in specially).

'So infinitesimal numbers exist.'

I have doubts. The method I have described for getting non-standard models of real number theory does not yield a *unique* model, and does not even yield models that are all isomorphic to one another. The same is true of other methods. So the expression 'the set of infinitesimal numbers' isn't well-defined, unlike the expressions 'the set of rational numbers' and 'the set of real numbers'. 'Infinitesimal numbers exist' doesn't, at present, have a clear sense.

Still, we do have:

> There are several sets of things that behave in the ways mathematicians have wanted infinitesimals to behave.

But was that existential proposition established *simply* by establishing consistency? I think not. At a crucial point you have to show that every finite subset of the axioms of R* *has a model*. Or else somewhere in proving the Compactness Theorem you have to appeal to a previously proved theorem that certain consistent sets of formulas *have models*. That is, at some point you have to establish that a *consistent* set of formulas *applies* to something.

Finally, there is one explicit result in logic where *formal* consistency, at least, *doesn't* entail having a model: for higher-order languages there are sets of sentences that are consistent but don't have any models (see e.g. Henkin *Trans. Amer. Math. Soc.* 74, 1953, p. 424).

University College of North Wales,
Bangor, Gwynedd LL57 2DG

© GEOFFREY HUNTER 1988

'EXIST' AND TWO TYPES OF QUANTIFICATION

By D. H. Rice

STRAWSON distinguishes two uses of quantifying adjectives ('Is Existence Never a Predicate?', *Freedom and Resentment*, pp. 189-97). Using them the first way we say, '*with regard to the members of a class the existence of members of which is presupposed*, that all, most, some etc. *of them* exhibit a certain characteristic.' (Call this 'type 1 quantification'.) The second use is to indicate (in the case of simple assertions) roughly how big a non-presupposed class is affirmed to be. (Call this 'type 2'.) Type 1 is exemplified by 'Some tame tigers growl', and type 2 by 'There are some tame tigers'; also by 'Some tame tigers exist' – at least in its natural interpretation. Whereas 'many', 'some', 'a few', 'no' and 'at least one' may be used in both ways, 'all' and 'most', it seems, can be used only in the first way.

Strawson goes on to suggest that the 'essence of predication is the ascribing of something to an individual, or to some, none or all of a class, where the existence of the individual or of members of the class is presupposed (or has been *antecedently* affirmed or implied)', and that for this reason we may want to say that 'exists', as it appears in 'Some tame tigers exist', is not used predicatively. He suggests further that there are other contexts in which 'exists' may be used predicatively – where, in particular, the presupposed class has fictional members (say) as well as real ones.

In what follows I shall not be concerned with this last point, about quantification over 'radically heterogeneous' classes. Indeed I shall assume that we have only (attempted) reference to the real world (of space and time) to consider. Supposing, however, that the rest of what Strawson says is largely right, I want to make some suggestions about the use of verb phrases with type 2 quantifiers. I shall also draw some conclusions about 'exist'.

First we may note that, whereas some verb phrases can be used with both types of quantifier, others cannot, or so it seems. 'Live in that forest' can be used with both types, for instance, but 'are red' and 'have stripy tails' can be used, I suggest, only with type 1 quantifiers.

It is relatively easy to show that a verb phrase can be used with type 1 quantifiers. For example, 'Most tame tigers live in that forest' makes good sense; so does 'Some tame tigers live in that forest, but some do not'. So 'live in that forest' can be used, we may conclude, with type 1 quantifiers. It is not quite so easy to show that a verb phrase can or cannot be used with type 2. However, I propose three tests, in an order that reflects decreasing confidence.

(i) Can one reply to the assertion that there are no *x*'s with 'On the contrary, (some) *x*'s *F*'? If so '*F*' can be used with type 2 quantifiers, and otherwise not. The rationale for this test is that one would not expect to be able to reply to an assertion that *x*'s do not exist with something that *presupposed* – took for granted – that they did. (Try this test with 'ghosts' together with 'haunt this house', on the one hand, and 'are frightening', on the other. But what about 'frightened my mother'?)

(ii) Is '*x*'s *F*' ambiguous, meaning either the same as 'All *x*'s *F*' or the same as 'Some *x*'s *F*'? If so '*F*' can be used with both types, and otherwise not. The rationale is, I suggest, that where 'some *x*'s *F*' means the same as '*x*'s *F*' the 'some' is of type 2. Then, of course, we would not expect ambiguity in the case of '*F*' if '*F*' could not be used with type 2 quantifiers.

(iii) Can 'Some *x*'s *F*' be rephrased as 'There *F* some *x*'s'? This is a tricky test to operate because in modern English we do not any longer use this device for inverting (grammatical) subject and predicate with such a wide variety of verbs. Nor does the inversion perhaps always take so simple a form as my specification of the test suggests. However, we can certainly say 'There exist some tame tigers' as a variant of 'Some tame tigers exist'. Indeed it is simplest to suppose that this is synonymous with 'There are some tame tigers'. That is to say, the 'are' in the latter means *exist*; what we have there is an inversion of the old-fashioned 'Some tame tigers are'. I take it, also, that 'live in that forest' passes the test. So does 'are eating your sandwiches'; though 'There are some tame tigers eating your sandwiches' sounds better than 'There are eating your sandwiches some tame tigers'. 'Are red', on the other hand, fails. 'There are red some tame tigers' is impossible. Of course we can have 'There are some red tame tigers', but there the 'are' surely means *exists*; at any rate it is not the copula from 'Some tame tigers are red'. One might suspect, it is true, that the 'are' in 'There are some tame tigers eating your sandwiches' also means *exist*. Maybe that *is* why that version sounds better to modern ears. However, we can surely have 'There have some tame tigers eaten your sandwiches'. So the 'are' *could* be the 'are' from 'are eating'.

I note in passing that 'is (are) identical to itself (themselves)' fails these tests. So being identical with oneself (*inter alia*) is not the same thing as existing.

What is it about some verb phrases that allows them to be used with type 2 quantifiers? Perhaps first we should ask what the rationale for test (iii) might be. *Collins English Dictionary* implies that the inversion is possible only 'when the true subject is an indefinite or mass noun phrase'. If that were so we would expect a verb phrase to fail the test if it could be used only with type 1 quantifiers. But Samuel Foote tells us, 'and there were present the Picninnies, and the Joblillies, and the Garyulies, and the Grand Panjandrum himself...' It is true that Foote was *trying* to write nonsense, but evidently he was not wholly successful. So it seems that the dictionary is wrong.

A more tempting suggestion (for which I am grateful to Michael Woods and Lesley Brown) is that we take the 'there' seriously. What the successful verb phrases have in common is that they signify the occupying, in some mode, of space. One might say that, where spatial location is not explicitly mentioned, 'somewhere' is always understood in the case of such verb phrases. To say that Tigger is eating your sandwiches is to say that Tigger is eating your sandwishes *somewhere*. Now, of course, we can say that Tigger is red somewhere (on his ear, as it might be), or that he has a tail somewhere. But neither 'is red' nor 'has a tail' signify the occupying of space by what they are applied to. The 'somewhere' does not indicate in either case where *Tigger* is, nor the verb phrase how *he* is occupying space. Again we can say, 'Somewhere skies are blue'; but I take it that the 'somewhere' there goes with 'skies'.

But the occupying of space just is, for the sorts of objects we are concerned with, existence. So we may further say that what the verb phrases that figure in inversion have in common is that they signify modes of existence. And now we can add something that was perhaps already implicit in what was said about the rationale for test (i). We said that verb phrases that *failed* it did so because they could be used only with type 1 quantification; in which case the riposte would be taking for granted what had just been denied. We may add now that the verb phrases that *pass* the test do so because, not only do they not take for granted what has been denied, they actually affirm it.

Let us take stock:

(a) Other verb phrases than 'exist' can be used with type 2 quantifiers.

(b) These verb phrases signify modes of existence.

We may add:

(c) The fact that they can be used non-predicatively does not, of course, stop them being used predicatively. And we do not have to invoke fictional items, or such like, to allow for this.

(d) We are not obliged to say, when they are used non-predicatively, that they do not continue to signify the possession of properties. Unless we will not count relations to particulars as properties.

But by the same token we may continue — in the absence of further considerations to the contrary — to believe that 'exist' always signifies a property: that of occupying space, in the case of spatiotemporal objects. We will expect also that it *can* be used predicatively, even without invoking fictional objects. And it can. 'Most tame tigers exist' and 'All tame tigers exist' make perfectly good *sense*. It is just that they are not worth saying; any more than it is worth saying that most tame tigers are tame tigers,

A final speculation. Even where there is no explicit type 1 quantification or explicit reference, there is at least *implicit* reference in the shape of the implicit 'somewhere'. So even 'Tame tigers exist' has a logical subject – namely space – of which it is predicated that it is occupied, some of it, by tame tigers.

Christ Church,
Oxford OX1 1DP

© D. H. RICE 1988

ON AN ARGUMENT OF ARMSTRONG AGAINST NOMINALISM

By MICHAEL LEVIN

NOMINALISTS are inveterate paraphrasers. Since a major part of the Realist's argument is appeal to facts allegedly expressible only by sentences containing reference to universals, Nominalists must find sentences expressing these same facts – sentences intuitively equivalent to the Realist's – which contain reference to individuals only. David Armstrong has recently pressed an objection to one of Nominalism's seemingly least controversial paraphrases (D. M. Armstrong, *Nominalism and Realism*, Cambridge 1978, pp. 58-61).

Realists point to sentences like

(1) Red is a colour

as ostensibly about a universal, namely red or redness. Nominalists reply that (1) is equivalent to

(2) (x) (x is red $\supset$ x is coloured),

a sentence whose quantifiers range over individuals only. However, argues Armstrong, if (2) is equivalent to (1), then the undoubted truth

(3) (x) (x is red $\supset$ x is extended)

must be equivalent to the patent falsehood

(4) Red is an extension.

(1) cannot be equivalent to (2), therefore, and the Nominalist's general paraphrase recipe – taking '*X*ness is a (form of) *Y*ness' to be equivalent to 'All *X*'s are *Y*'s' – must also be wrong.

I doubt that the Nominalist can modify his general recipe so as to forbid intuitive inequivalences like that of (3) to (4), since I see no plausible principle on which such a modification might rest. It certainly will not do to capitalize on the non-analyticity of (3) by decreeing that only analytic 'All *X* are *Y*' statements entail '*X*ness is a (form of) *Y*ness' counterparts, or that '*X*ness is a (form of) *Y*ness' is well-formed – hence susceptible of paraphrase – only when analytic. (Judith Myers made both suggestions in conversation.) 'Manliness is a virtue' is certainly not analytic, yet it is well-formed and intuitively equivalent to 'All manly acts are virtuous'. Although I cannot prove this *a priori*, further efforts to isolate cases resistant to Nominalistic paraphrase seem likely to become ever more *ad hoc*.

So the Nominalist had better reconcile himself to inferring (4) from (3), and set about explaining away the *appearance* that (4) is false. This brave resolution defies all manner of intuitions, for instance that someone unfamiliar with colours (e.g. a blind man) could master the word 'extended', while mastery of 'colour' does require familiarity with some particular colour. At the same time, this cluster of epistemological and linguistic intuitions won't do much work for the Realist unless it reflects some underlying metaphysical intuition, and in fact it does – that redness, or being red, is a *determination* of the *determinable* 'being coloured', while redness is not a determination of the determinable 'being extended'. The issue is best joined, then, by construing statements of the form 'Xness is a (form of) Yness' in determinable/determination language. (1) says, plausibly enough, that being red is a way in which things can be coloured, a form of colouredness; what the Nominalist must make convincing is that being red is also a determination of the determinable 'being extended', a way in which a thing can be extended, a form of extendedness.

Uncomfortable with talk of universals, the Nominalist does not fully grasp expressions like 'a way of being Y'. He nevertheless appreciates that 'a way of being extended' is most naturally taken to refer to ordered pairs of sizes and shapes. Being a cube eight cubic feet in volume is the sort of thing that should determine 'being extended', while being red is best taken as 'a way of being coloured'. On closer inspection, however, being red *can* determine extendedness as well as colouredness. To the extent that the Nominalist follows such talk, $X_1, \ldots, X_\alpha$ are the determinations of the determinable Y just in case any actual or possible Y thing is (or would be) an X_1 thin or ... or an X_α thing. In this sense colours are indeed determinations of extendedness, since, ignoring irrelevant complications created by transparent panes of glass, any extended thing (or its surface) is red or green or.... All the red sized shapes and all the green sized shapes ... yield all the sized shapes there could be. (4) is thus true and Armstrong loses his counterexample.

To be sure, a determinable may be more picturesquely conceived as a spectrum whose parts or stretches are its determinations, with the colour spectrum composed of particular colours and the extension spectrum composed of size-shapes under a more complex ordering. Since the Nominalist would be giving away the game were he to regard the extension spectrum (or any other spectrum) as constructed from universal size-shapes, he takes it as constructed from what have been called particularized universals – the size-shapes of extended particulars. The Nominalist now claims that the spectrum of extensions can be constructed from segments composed of different colours if all extensions are taken to be the colours they actually are (or would be if they existed). If all coloured things are necessarily extended and all extended things necessarily some colour, being red is a more specific way in which a thing can be extended just as surely as being a cube of volume eight cubic feet is. The red extensions form a gerrymandered stretch of the extension spectrum under its natural ordering, just as the cubes between eight and ten cubic feet in volume form a natural stretch. I suspect, in fact, that red seems not to be a determination of extension only because of the different roles of red in the topologies of the colour and extendedness spectra. Not only is red a determination of 'being coloured', any determination of 'being coloured' located between two reds is itself a red. Under the natural ordering of extensions, however, there are pairs of red determinations of extendedness that surround determinations of extendedness that are not red. A violet cube of volume nine cubic feet might, for instance, separate two cubes of volumes eight and ten cubic feet that are different shades of red. Extensions do not fall into dense segments defined by particular colours, while colours *do* break up into dense segments defined by particular colours.

The obvious objection to specifying the sized shapes in the world as the red ones, the blue ones, the green ones ... is that there are red *and green* cubes eight cubic feet in volume. The colour of the extension 'cube eight cubic feet in volume', a unique position on the extendedness spectrum, becomes indeterminate. The red and green volumes overlap improperly. In contrast, every position on the colour spectrum is precisely one shade, and every position on the extension spectrum is precisely one size-shape. Colours, therefore, cannot determine extendedness in the way colours determine colouredness and size-shapes determine extendedness. What this asymmetry shows, in turn, is that the extension spectrum is composed of specific extensions conceived as universals, not of extended particulars or the particularized extensions of particulars, just as the colour spectrum is composed of specific colour universals, not coloured particulars or the particularized colours of particulars. And what *this* shows, the objection concludes, is that talk of determinables and determinations *is* inherently Realistic, and untranslatable into a Nominalistic ersatz — just what (4) was intended to show.

Technical expedients meet the first part of this objection. The colours of representatives of each lot of congruent particulars can represent particular size-shapes, thus preventing improper overlap of the red and green extensions. If necessary, each representative can be assigned a different colour, so that to each absolutely specific extendedness universal there corresponds an absolutely specific colour. But what of the intuitive point that (2) and (4) are discrepant because the colour spectrum is made of colours and the extension spectrum is made of extensions? This is an intuition that the Nominalist has no obligation to capture. The claim that Nominalism entails a falsehood in the form of (4) is met by translating (4) into a Nominalistic truth not too removed in intuitive import from (4). It begs the question to complain that (4) is nonetheless false in the sense *in which it is about universals*. The ineliminability of reference to universals in '*X*ness is a (form of) *Y*ness' sentences is supposed to follow from their untranslatability into Nominalistic language, which is in turn supposed to follow from the discrepant truth values of (3) and (4). It is thus circular to base the allegation of discrepancy on the claim that (4) is about universals (while (1) is about different universals and (3) is not about universals). Nominalists owe the world an account of '*X*ness is a (form of) *Y*ness' sentences, since some of them are true. It is not clear, however, that such sentences are ordinarily regarded as true in the sense in which they entail the existence of universals. By providing an 'unnatural' true interpretation of (4), the Nominalist shows the truth of (4) not to be tantamount to recognition of universals. There may be senses of '*X*ness is a (form of) *Y*ness' in which its truth *is* tantamount to recognition of universals, but Nominalism may repudiate these senses without scandalizing common sense.

A methodological moral emerges: don't ask the Nominalist to translate scarcely disguised declarations of Realism. There is a tendency among Realists to think they can stump the Nominalist by picking some characteristic C from among the many putative characteristics of universals found objectionable by Nominalists, and then daring the Nominalist to render '$\exists xCx$'. A well-known Realist once challenged me to produce a Nominalist translation of 'There are uninstantiated properties'. Well *of course* the Nominalist can't translate *that* into a Nominalistic truth. He regards it as false or meaningless. All the Nominalist need preserve are sentences that you don't have to be a Realist to accept, the artificiality of the preservation being immaterial. Nominalists can of course see the point of saying that redness is not a form of extendedness *in the way* that being a cube eight cubic feet in volume is a form of extendedness. But since Nominalists do not think this saying is literally true, they need make nothing of it. They need only make something of more innocent statements, like (1) and (4).

City College of New York,
New York, NY 10031, U.S.A.

© MICHAEL LEVIN 1988

LEWIS'S MODAL REALISM: A REPLY TO NAYLOR

By MARK F. SHARLOW

IN a recent note ('A Note on David Lewis's Realism about Possible Worlds', ANALYSIS 46.1, January 1986, pp. 28-9), Margery Bedford Naylor asked why, if we accept David Lewis's classic argument for the reality of possible worlds (in *Counterfactuals*, Blackwell 1973, p. 84), we should not accept a seemingly similar argument for the reality of *impossible* worlds. The argument Naylor offered was a variant of Lewis's obtained by replacement of possibility terms ('could', 'might', 'possible') with corresponding impossibility terms ('could not', 'might not', 'impossible'). In the present note I will attempt to show that Naylor's argument has no force for an adherent of Lewis's position.

First I will summarize the two arguments. Lewis's argument begins with the 'uncontroversially true' premise

(L1) Things might have been otherwise than they are,

and proceeds to

(L2) Things could have been different in countless ways

and finally to a 'permissible' paraphrase of (L2),

(L3) There are many ways things could have been besides the way they actually are.

Taking (L3) 'at face value', Lewis concludes that there are 'ways things could have been'; he calls these ways 'possible worlds'. In Naylor's argument, the premise

(N1) Some things might *not* have been otherwise than they are

is used to infer

(N2) Things could *not* have been different in countless ways,

which admits the paraphrase

(N3) There are many ways things could *not* have been besides the way they are.

These 'ways' Naylor calls 'impossible worlds'.

Suppose — just for the sake of argument — that Lewis's modal realism is correct. Then we can formulate truth conditions for modal sentences using possible worlds in familiar ways; for example, 'Possibly *P*' is true iff *P* is true at some possible world accessible to the actual world. We can evaluate the validity of Naylor's argument from a modal realist point of view by applying such truth conditions to (N1)-(N3). (Assuming the existence of possible worlds to criticize a critic of such worlds would be question-begging, but I am not doing this here. I am constructing a possible worlds argument only to find out what an adherent of Lewis should think about Naylor's argument.)

(N1), 'Some things might *not* have been otherwise than they are', is true iff there is some feature of the actual world which is not lacked by any world accessible to the actual world. Equivalently, (N1) is true iff there is some proposition true at the actual world and at all possible worlds accessible to the actual world. (Such a proposition is just a necessary proposition.)

(N2), 'Things could *not* have been different in countless ways', is harder to understand than (N1). (N2) could mean that there are many impossibilities ('things could not have been such that *p*, or such that *q*, or...') or that there are many necessities ('things could not have failed to be such that *p*, and such that *q*, and...'). Naylor's example of a chair that cannot be two different colours in the same respect simultaneously suggests to me that Naylor intended the former meaning. Fortunately, this ambiguity need not sidetrack our analysis; we can show that the truth conditions come out the same on either reading. According to a standard law of modality negation, it is impossible that *p* iff it is necessary that not-*p*. (For example, it is impossible for Naylor's chair to be two colours in the same respect simultaneously if and only if it is necessary that the chair is not that way.) Similarly, it is impossible that not-*p* iff it is necessary *p*. Thus out of each pair {*p*, not-*p*} of propositions, one member is impossible iff the other is necessary. Since the relation between propositions and their negations is one-to-one onto, there is a one-to-one correspondence between necessities and impossibilities. Hence there are many impossibilities iff there are many necessities, and the two readings of (N2) are materially equivalent. (N2) is true iff there are countless necessary propositions – or equivalently, many features of the actual world not lacked by any world accessible thereto. Strictly speaking, the step from (N1) to (N2) is invalid because of the countlessness condition in (N2). However, this flaw is unimportant because it may be repaired without violence to the intent of the argument – for example, by replacing 'some' with 'countless' in (N1).

(N3), taken at face value as Naylor takes it, differs from (N2) and (N1) in its explicit mention of 'ways' or worlds. Taken this way, (N3) is an existential statement which is true iff there are many impossible worlds. But (N1) and (N2) have no existential consequences for possible or impossible worlds. They are true at the actual world no matter which other possible worlds exist; they don't require the existence of any impossible worlds for their truth. Thus (N3) is not a consequence of (N1) or (N2). The inference from (N1) and (N2) to (N3) is invalid.

This argument shows that an adherent of Lewis need not take Naylor's argument seriously as a criticism of Lewis, or for that matter as an invitation to believe in impossible worlds. Such an adherent is committed to modal truth conditions according to which Naylor's argument does not preserve truth; to him or her, Naylor's argument is invalid. On the modal realist account of possibility, the existence of impossibilities can be explained using possible worlds alone, so the existence of impossibilities does not require the existence of impossible worlds.[1]

P.O. Box 2039, Santa Monica,
California 90406, U.S.A.

© MARK F. SHARLOW 1987

---

[1] I wish to thank Christopher Kirwan for his valuable comments on an earlier version of this paper.

A NOTE ON OBJECTS AND EVENTS

By OLAV GJELSVIK

THERE are two main traditions among those who hold that events are particulars in space and time. One view, which we might call Strawsonian, holds that the category of substance is conceptually prior to the category of event (Strawson 1959, pp. 46–7). The other view, which we might call Davidsonian, holds that there is no conceptual priority either way between these two categories, but rather an underlying symmetry of conceptual dependence. Still there is no denying that we very often identify events by identifying a substance undergoing a change (Davidson 1980, pp. 173–5).

Writers in the Strawsonian tradition find it natural to give identity-criteria for events in terms of substances exemplifying properties at times. Events are identical when their constitutive substances, properties and times are. This account of identity-conditions for events might be preferred even if one does not hold the priority-thesis, but if one does hold it, one has a good motivation for exhibiting this priority in the theoretical account of what makes an event the event it is. J. Kim has in a sequence of papers given details of such a view (Kim 1969, 1973 and 1976). His theory faces a number of problems, and the debate around it has focused on what is particularly urgent: to be able to draw a principled distinction between properties which are constitutive of (or essential to) events and properties which are not.

There is a different problem for the Kimian approach which has been given little notice, and which deserves some attention. The problem is this. We rotate a little statue slowly. The debate referred to above has discussed the question whether ⟨rotating⟩ or ⟨rotating slowly⟩ is the constitutive property of this event. But a distinct question is this: what is rotating? What is the subject of this change? The statue, Kimians are likely to reply. But this reply is not obviously correct. The reason is this. There are very good grounds for thinking that there can be more than one object at the same place at the same time. In this case we should hold, I think, that the statue is not identical with the piece of bronze it is made of, even if they have the same shape, weight and causal properties. Of course both the statue and the piece of bronze are rotating. But since they are not identical, it follows, given Kim's criterion, that the rotating of the statue and the rotating of the piece of bronze are not one and the same event, since they have different constitutive objects. This seems counter-intuitive and perhaps unacceptable. We think of the rotation as one event, and it is hard to swallow that there are two rotations, one of the statue, and one of the piece of bronze. We should also note that our unwillingness to think that there are two rotations can be nicely explained by the Davidsonian view which individuates events by causal considerations. Since these supposedly distinct rotations have exactly the same causes and exactly the same effects, there are no good reasons for thinking that there are two rotations: there is one rotation, individuated by its causal role.

I think one should avoid the consequence that there are two rotations, and if one agrees with this, then there are two possible options if one wants to stay within the Kimian tradition. The first is to modify Kim's criterion in order to make such cases impossible. The simplest way of doing this is to give identity-criteria for events by their spatio-temporal location and property exemplified. In the present case we will, of course, identify the spatial location by identifying the spatial location of the objects which exemplify the property of ⟨rotating⟩ at the time in question. But there is one difficulty with this line, and that is that one seems unable to exhibit the conceptual priority of the category of substance if identity-criteria for events are given by spatio-temporal locations and properties exemplified at those locations. A different version of this response is to lay down as an axiom that there is only one event when objects which occupy the same place at the same time exemplify the same property at a given time. This also calls for an explanation and a justification, and it is hard to see how such an explanation can be entirely in accord with the attempt to exhibit the priority of substances in the theory of event-identity, since it is most natural to introduce causal consideration about events in such an explanation.

The other option is this. One does not modify Kim's definition, but sticks to the view that there is only one object which can be the constitutive object of an event at a place at a time. This can be motivated in different ways. One can attempt to defend the view that there is only one object at a place at a time, but this does not look very promising in the light of the good arguments against the view that there is identity between the statue and the piece of bronze it is made of. Another line would be to argue that the piece of bronze is not the kind of object which can be the subject of an event, or a subject of change. This seems to be the line taken by L. B. Lombard in his recent book *Events, A Metaphysical Study* (Lombard 1986, p. 70). But I see no good reasons for the claim that pieces of bronze cannot be the subjects of changes or events. If a piece of bronze falls down, that is an event. If this line is to be plausible, we must have a well motivated principle to distinguish between objects which can be subjects of events and objects which cannot. This principle must make it impossible that there is more than one subject of change at a place at a time, explain why this is so, and individuate events so that most of the events we recognize as events are counted as events. I have not seen such a motivated principle. Lombard does not give a positive criterion for what a subject of change might be. His examples are the main source for what his view is, but these give little promise: he counts a snowflake as a subject of change, but not a piece of bronze. I do not see how a well motivated principle can distinguish between a snowflake and a piece of bronze in this respect. At present I will leave the task of supplying a principle to those who want to defend this line; in my mind this line does not seem much more promising than the line which holds that there is only one object at a place at a time. If one were to approach the mind-body problem with a view like Lombard's, one would have to deny either that persons or that their bodies are subjects of change (or events). That does not look like a promising or interesting way of approaching this problem.

I conclude that the reasoning around this little problem exhibits a difficult problem for the Kimian approach as it is stated. Awareness of the problem must most likely result in some revision of Kim's criterion of event-identity. Such a revision might easily make the revised theory less palatable for someone who holds the Strawsonian priority-thesis. If one wants to avoid such a revision, one must supplement the account with a new principled distinction, and this might be very hard to do. As a whole I think the point gives some support to a causal individuation of events and thereby the Davidsonian non-priority view. The details of a causal individuation of events, however, remain to be given. Perhaps sameness in causal role should not generally be regarded as sufficient for event-identity, even if it can be thought generally necessary, and sufficient in the case considered in the text.

University of Oslo,
Box 1024 Blindern, 0315 Oslo 3, Norway

© OLAV GJELSVIK 1988

REFERENCES

Donald Davidson, *Essays on Actions and Events*, Clarendon Press, Oxford, 1980.

Jaegwon Kim, 'Events and Their Descriptions', in Nicholas Rescher (ed.), *Essays in Honor of Carl Hempel*, D. Reidel, Dordrecht, 1969, pp. 197–215.

Jaegwon Kim, 'Causation, Nomic Subsumption, and the Concept of Event', *Journal of Philosophy*, 72 (1973), 217–36.

Jaegwon Kim, 'Events as Property Exemplification', in Myles Brand and Douglas Walton (eds.), *Action Theory*, D. Reidel, Dordrecht, 1976, pp. 159–77.

Lawrence Brian Lombard, *Events. A Metaphysical Study*, Routledge and Kegan Paul, London, 1986.

Peter F. Strawson, *Individuals*, Methuen, London, 1959.

THE PRAGMATIC PROBLEM OF INDUCTION

By JOHN WATKINS

I WISH to present a simpler and, I think, better version of the solution of the pragmatic problem of induction put forward in my *Science and Scepticism* (London: Hutchinson, pp. 342–8).

Let a technologist confront a practical problem upon which several rival hypotheses bear, each indicating a different solution. One of these is better corroborated than its rivals. Can a neo-Popperian like myself, who seeks to exclude all non-deductive inferences and who insists that corroboration appraisals of hypotheses have no implications for their future performance, give him any reason why he should prefer the best corroborated hypothesis?

My solution relies on two assumptions. One is the historical assumption that well corroborated hypotheses *have hitherto* proved better guides for practical decision-makers than have hypotheses that were not well corroborated. After all, the Humean problem of induction is not about the reliability of our knowledge of the past. Hume did not challenge the proposition that it *has* paid better to be guided in one's actions by a principle of induction than by some alternative principle. His problem was: is there any reason to act on the supposition that it will continue to pay better in the future?

My other assumption concerns rational choice under conditions of uncertainty. Suppose that you are planning a holiday outing for tomorrow. You are in a district where the weather is rather variable and unpredictable. You are considering two alternatives, A and B. For A to be a success, only one weather condition must be satisfied: it mustn't be windy. For B to be a success, it must be neither windy nor rainy. Either outing will be equally enjoyable if not spoilt by the weather, and equally unenjoyable if so spoilt. Then it would clearly be rational for you to choose A, since the success of B requires the fulfilment of two conditions, say C_1 and C_2, whereas A only requires that of C_1.

Now suppose that you have a choice between *A* and an equally enjoyable/unenjoyable alternative B', and that for B' to be a success it mustn't rain but it must be windy (perhaps it's a surfboarding expedition). In this district, with its highly unpredictable weather, windy days are about as frequent as non-windy ones. Then it would again be rational for you to choose *A*, since the success of B', like that of the previous *B*, requires the fulfilment of two conditions, not-C_1 and C_2, whereas *A* only requires that of C_1, and you have no reason to regard C_1 as less likely than not-C_1.

To put it more generally, it is rational, other things being equal, to choose that course whose success presupposes *the weaker forecast about the future*. For a weaker forecast is less likely, other things being equal, to turn out wrong than a stronger one.

Now consider two rival principles. One says that, during the whole history (past and future) of mankind, the overall success-rate of actions guided by well corroborated hypotheses is higher than that of actions guided by hypotheses that are not well corroborated. We may call this a *corroborationist principle*. The other, which we may call a *counter-corroborationist principle*, has two versions. In the first, it says that, during the whole history of mankind, the overall success-rate of actions guided by hypotheses that are not well corroborated is higher than that of actions guided by well corroborated hypotheses. This version is simply the contrary of the corroborationist principle. In the second version, it says that during the *future* history of mankind the overall success-rate of actions guided by hypotheses that are not well corroborated will be higher than that of actions guided by well corroborated hypotheses. This version is a 'grue-ish' variant of the corroborationist principle.

Denote by *E* our historical assumption about the higher success-rate of actions guided by well corroborated hypotheses in the past. Either version of the counter-corroborationist principle, when conjoined with *E*, *yields a stronger forecast about the future* than that yielded by the corroborationist principle when conjoined with *E*. In its first version, the counter-corroborationist principle implies, not merely that the success-rate of actions guided by hypotheses that are not well corroborated will in the future be higher than that of actions guided by well corroborated hypotheses, but that it will be *sufficiently higher* to offset its lower success-rate in the past. No counterpart of this is yielded by the corroborationist principle when conjoined with *E*. In its second version, the counter-corroborationist principle, when conjoined with *E*, implies that a time will come when success-rates change over and, moreover, that *this time has come now*. Again the corroborationist principle says nothing corresponding to this when conjoined with *E*.

Thus our two assumptions jointly imply that it will be rational to allow the corroborationist principle, rather than a counter-corroborationist principle, to guide our actions, since it involves the weaker forecast about the future. Are these two assumptions in order?

The second seems ungainsayable. But some critics of my book have challenged the first: why should we accept E?

I do not claim that we know E to be true, but I do say that unless we accept E, or something like it, the problem of induction would not even arise. It can arise only if it is allowed that there is *something* from which inductions *might* proceed. The question then is whether they *should* proceed. Hume took it for granted that certain regularities have obtained in the past, and spoke confidently of the observation of constant conjunctions. A critic might have taken him up on this. Shouldn't he have been more sceptical here? Do we really know that every raven so far observed has been black? Do we really know that all emeralds observed before t_0 have been green? Can we be sure that the laws of nature have held sway down to the present?

Hume would no doubt have conceded that we have no perfect assurance about such matters. But he might have added that *if* we suppose ourselves to have some reliable knowledge about the past, then the interesting problem arises: would we be entitled to make any inductive use of that knowledge? In defence of my introduction of E, I say that someone who proposes a solution of the problem of induction is entitled to introduce an assumption that is needed for the posing of the problem.

The problem of induction has seemed intractable because we cannot, without circularity, make any *inductive* use of any historical assumption about the past, such as that denoted by E; from which it has generally been concluded that we cannot make *any* use of it. The trick of the present solution is to use E but in a *non-inductive* way.

The London School of Economics,
Houghton Street,
London WC2A 2AE

© JOHN WATKINS 1988

FEAR OF SPHEXISHNESS

By RICHARD DOUBLE

IN his extremely intriguing *Elbow Room*, Daniel Dennett tries to show how human dignity, freedom, and moral responsibility can exist in a materialistic universe lacking in libertarian, contra-causal freedom (Dennett, 1984). A large part of the task Dennett sets himself is to debunk traditional incompatibilist 'bugbears' designed to 'pump' our intuitions that any sort of determinism is as destructive to human freedom as discovering we are controlled by the Invisible Jailer, the Nefarious Neurosurgeon, the Hideous Hypnotist, etc. One of these spectres arises from a fear of 'mechanism' illustrated by the digger wasp (Sphex ichneumoneus). In order to provide undecayed food for her offspring when they hatch, Sphex paralyses a cricket and then brings it inside her burrow next to her eggs before sealing the burrow and leaving. In bringing the live food home Sphex always follows a seemingly reasonable strategy. Sphex leaves the cricket at the threshold of the burrow, goes inside alone ('to see that all is well'), and then drags the cricket inside. Sphex's stupidity is revealed when a biologist moves the cricket a few inches back from the threshold while Sphex is inside. Sphex again pulls the cricket up to the threshold and goes back alone inside the burrow. Sphex will engage in this stereotyped behaviour all day long, never figuring out that she can pull the cricket into the burrow from the place inches distant from the threshold (11).

Dennett imagines an incompatibilist 'fearmonger' saying

> What makes you so sure you're not sphexish — at least a little bit? ... Could a superior intelligence, looking down on us, find a similar if more sophisticated trick that would unmask us? (p. 11)

Dennett takes the threat posed by the Sphex example to be worthy of serious discussion, since (i) there are some of us who are importantly sphexish (the insane, retarded, and brain-damaged), and (ii) some recent psychological literature shows most of us to be more sphexish than we realize (Milgrim's obedient torturers, the rationality literature, and our susceptibility to operant conditioning of which we are unaware) (pp. 12-13).

But what, exactly, is the fear of sphexishness that Dennett wishes to dispel? On page 10 Dennett says that the worry Sphex raises is that human deliberation might be 'mechanical', that we might 'not be free agents, but only *automata*'. One page later Dennett claims that the Sphex fear is that we could be revealed as having a 'mindless mechanicity' similar to that of Sphex. On the face of it, these are two conceptually different problems, since it is easier to imagine being shown that one's behaviour had been under the control of forces one did not recognize than it is to imagine being shown that one had been acting and deciding without intelligence. Thus, I think we should distinguish two fears of sphexishness:

(FS1) Someday, someone knowing enough about the determinants of our behaviour could control us as completely as we control Sphex.

(FS2) Someday, someone knowing enough about the determinants of our behaviour could show that we are just stupid syntactic engines like Sphex.

Let me state my disagreement with Dennett concerning these fears. I endorse Dennett's announced compatibilist line that deter-

P4462

minism *per se* is no threat to human freedom, agreeing that 'what makes the wasp's fate so dreadful is not that her actions and "decisions" are *caused* but precisely that they are so *simply* caused' (p. 12). I agree also that (FS1) is not a rational worry, since we are not likely to meet such a being (p. 61). I think Dennett fumbles the compatibilist ball, however, with respect to (FS2), since he actually gives an argument for the view that we ultimately are sphexish in that sense. In what follows, I shall consider that argument of Dennett and propose an alternative to show that the compatibilist has a stronger reply to (FS2) than merely that we will not meet such a being.

I think that Dennett's endorsement of ultimate sphexishness is due to two arguments. The more general argument is that our biological heritage puts us on the same slippery slope as Sphex, thus making us vulnerable to the same failings. For Dennett interests, reasons, and the ability to appreciate reasons are all the products of the evolutionary past of all organisms, so whatever epistemic limitations infect the lower phyla such as Sphex, in principle will infect humans also (pp. 22-6).

By itself, I doubt whether this slippery slope argument would impress many readers, but Dennett supplements it with what might be called a 'Syntax to Semantics' argument:

(i) Brains are merely syntactic engines:

> brains are just very complicated physical organs; whatever they react to must be some physical change or difference in the stimuli they encounter ... as physical mechanisms they can only be *syntactic engines*, responding only to structural or formal properties (p. 28)

(ii) Syntax does not entail semantics:

> how does the brain manage to get semantics from syntax? It couldn't. Syntax all by itself doesn't determine semantics. It is a relatively straightforward task to design and build machines that test for acidity. ... But think of designing a machine that tests for the presence of hostility or love or danger or skepticism. ... Since meaning does not reside, like some rare ore, in physical features of stimuli, no alchemical extraction process could distill it and respond to it (p. 28)

(iii) Therefore, brains do not really ascertain semantics, they only seem to:

> brains ... are not semantic engines ... brains only approximate the behavior of the (ideal, pure) semantic engine. ... They are cunningly designed to be not infinitely but indefinitely sensitive to meaningful changes, not perfectly but practically invulnerable to sphexishness (p. 29)

Whether one sees this as an *argument* or something looser, it seems clear that to be persuasive it needs at least the additional claim that the only possible evidential link between syntax and semantics is a logical one, a claim reminiscent of those made by neo-Wittgensteinians concerning other minds (Chihara and Fodor, 1965). To help us envision an alternative to Dennett's view, let us consider a parallel argument concerning other minds:

(i*) Humans are able to respond only to the observable physical properties of other persons.
(ii*) The observable physical properties of other persons do not entail that they have minds.
(iii*) Therefore, humans do not really ascertain that other persons have minds, we only seem to.

If you do not like the latter argument, you should not like Dennett's either. Viewing the other minds argument suggests, I think, that the best strategy for justifying our belief in other minds is also the best strategy for resisting (FS2), *viz.* inference to the best explanation. Let me briefly sketch how.

The simplist capacity of an organism's nervous system is that of detector of physical properties: if a nervous system (or CPU) has only this capacity, it will be merely a syntactic engine. For instance, Sphex's brain does not assign any meaning to the fact that it has just checked out the burrow five seconds ago – Sphex 'knows' only that it is to check the burrow after it drags the cricket to the threshold, irrespective of how many times it has done so before. Likewise my IBM PC does not understand the futility of the infinite loop that my program instructs it to follow. The 10,000th time it reads my instruction 'Goto 500', my instruction still means to execute the instruction on line 500 roughly in the way a kick in the pants means 'Here, have a little pain'.

A second level of complexity arises in organisms that are able to take account of Gricean natural meaning (Grice, 1957). Here organisms take stimuli as evidence for the occurrence of further stimuli without imputing an intentional agent who *means* anything by the stimuli, as in Grice's example that spots on a person 'mean' measles. At the simplest end this recognition of natural meaning might be extremely rigid, shading into level one; we might say Sphex 'takes' the wasp's presence on the threshold as 'evidence' that now it is time to go into the burrow and look around. Yet in creatures possessing greater capacities of memory and concept formation, the recognition of natural meaning might lead to fairly plastic behaviour as when Rover is alarmed by the tone of master's voice or, to use a favourite example of Dennett, when vervet monkeys take a characteristic cry from other vervets as evidence that an eagle is about (Seyfarth and Cheney, 1980).

The syntactic engine becomes a semantic one at the third level where organisms are able to impute Gricean non-natural meaning to syntax. Here brains take certain physical stimuli to be products of the intentions of minded entities to produce them with the intention that they be so recognized. What would a brain (or a computer, for that matter) have to be like to manage this? It seems that it would have to be able to represent to itself various hypotheses about what explains those stimuli and assign probabilities to these. Thus brains would need a certain range of concepts – causality, agency, and intention for openers. It may well be that to have a concept of agency to apply to a sign producer, the brain will need a comparably sophisticated concept of agency in the first person. Specific concepts regarding the particular hypothesis at stake will be needed also. In sum, the syntactic engine becomes (phylogenetically or ontogenetically) a semantic engine when it can do with its own syntactic input what persons do with respect to their data on other minds. (Note: none of this seems to require conscious rehearsing; indeed there is every reason to believe it is unconscious.)

Perhaps this sketch sheds some light on our bias against machine intelligence. As Dennett notes, the mind boggles at the idea of building a machine that tests for hostility or love (p. 28). On the account above, determination of these qualities requires an inference to the best explanation of observed phenomena (level three), while current digital computers respond solely to syntax (level one). This, however, shows nothing about the ultimate capacities of machines, even of machines defined solely in terms of the programs they run rather than their causal powers (cf. Searle, 1980). The fact that present machines are exploited as mere syntactic engines does not even seem to be the right sort of evidence for proving that they could not perform the third level imputation of causes required to turn them into semantic engines, nor, *pace* Searle, that they cannot obtain this ability solely as a result of their symbol manipulational capacities. As sketched above, the inference to the best explanation will need more sophisticated concepts than those needed by syntactic engines, and perhaps these cannot be gotten by a computer. But perhaps they could. Either way, it is easy to see that one argument against machine intelligence is very close to question-begging: a machine could impute semantics to syntax only if it had a concept of agency in its own case, which it cannot have since it is not an agent. The problem with this line is that it seems to entail that children can never learn to impute meaning to the signs they encounter. Perhaps by a gradual interactive process a computer's self-image as an intentional being and its ability to impute semanticity to symbols might both develop.[1]

In light of this positive account let us consider Dennett's position on (FS1) and (FS2). Dennett claims that it is logically possible that a being possessing Laplacian super intelligence would be as capable of controlling the outcome of roulette as an expert billiards player can control what happens on a pool table (p. 61). (Actually, this assumes *skill* ('know-how') as well as propositional knowledge – pool seems inherently more controllable than dropping a roulette ball 'just so'.) 'Such a super human intelligence could also, no doubt, control one of us as easily as we control Sphex' (p. 61). We can now see why this claim is misleading. We must grant that, on determinism, a superhuman intelligence with complete knowledge of us and the world could predict our thoughts and actions (assuming that the being did not interact with us). And if that being had also the skill to manipulate the determinants of our behaviour, it could determine our behaviour as we can that of Sphex.

---

[1] If the above sketch is on the right track, the process of inference would seem to be robustly mental, not in that it would show that semantic engines possess nonphysical properties, but in the sense that semantic engines are not, as Sphex perhaps is, a behaviourist's stimulus-response machine. Hypothesis generators seem to possess what Searle calls 'intrinsic intentionality', which is more fundamental than Dennett's notion of having the 'intentional stance' taken towards them.

All this shows only that (FS1) is possible, but it has no such force with respect to (FS2). Moreover, consider what such a being would have to do to exercise control over us. It would not suffice simply to relocate parts of our environment, as we do with Sphex's, since we, unlike Sphex, are semantic engines. To control us without our knowledge, the super genius would have to short-circuit our reasoning process altogether. Now comes a dilemma. If the super genius left our ability to generate and select among hypotheses intact, but scrambled our reasoning process so as to control us without our knowledge, then we would not be shown ultimately sphexish. Sphex's plight is pathetic; in this scenario ours would be tragic. If the super genius obliterated our capacities as hypothesis generators, then we would indeed become sphexish, but only because we have had our natures changed, not because we had been sphexish all along.

The morals of this discussion are these. (1) We should reject the epistemic boundedness problem as Dennett draws it. It is not that we display admirable, though ersatz, rationality up to a point and then, all at once, reveal our ultimate clockishness. Instead, although we are imperfectly rational agents throughout, since we are semantic engines early on, there is no threshold beyond which our intelligence would vanish. (2) Since we are not ultimately sphexish in my second sense, we do not need to endorse 'our policy of *not looking too closely* when looking at ourselves and each other as agents' (p. 72). *Some* causal processes are destructive of our freedom, of course, but the belief that causation in general has this result strikes me as startlingly close to incompatibilism, Dennett's announced foe. (3) If the imputation of semanticity can be understood as an inferential process which is biologically produced (the Grice/Searle line), then a common theme that unites the philosophies of language and mind may be applicable to the free will discussion, a theme which avoids both Cartesianism and behaviourism.

Indiana University-Purdue University,
2101 Coliseum Boulevard East,
Fort Wayne, Indiana 46805, U.S.A.

© RICHARD DOUBLE 1988

REFERENCES

C. Chihara and J. Foder, 'Operationalism and Ordinary Language: A Critique of Wittgenstein', *American Philosophical Quarterly* (1965), 281-95.

D. Dennett, *Elbow Room*, Cambridge, Mass., MIT Press, 1984.

P. Grice, 'Meaning', *The Philosophical Review* (1957), 377-88.

J. Searle, 'Minds, Brains, and Programs', *The Behavioral and Brain Sciences* (1980), 417-24.

R. Seyfarth and D. Cheney, 'Monkey Responses to Three Different Alarm Calls', *Science* (1980), 801-3.

SEARLE ON THE BIOLOGICAL BASIS OF COGNITION

By ASHER SEIDEL

SOME portion of John Searle's polemic against what he terms 'strong artificial intelligence' is based on his claim that any system possessing human cognitive abilities would need to duplicate the causal powders of the human brain. There is an apparent implication, drawn from this claim and further premises, that such duplication by means of nonbiological computational machinery is unlikely. This implication will be examined here.

In the 'Author's Response' section of his article 'Minds, brains, and programs' (*Behavioral and Brain Sciences* 1980, 3), Searle states:

> ... the strong AI [artificial intelligence] thesis ... implies that of all known types of specifically biological processes, from mitosis and meiosis to photosynthesis, digestion, lactation, and the secretion of auxin, one and only one type is completely independent of the biochemistry of its origins, and that one is cognition. The reason it is independent is that cognition consists entirely of computational processes, and since these processes are purely formal, any substance whatever that is capable of instantiating the formalism is capable of cognition. (p. 450)

In the above quotation Searle is making the following two claims. First, that the strong AI thesis makes a unique exception of human cognition in relation to all other biological processes, in that only cognition is considered completely separable from its biochemical basis. Second, that the reason for the first claim, as given by the strong AI thesis, is that cognition can be understood as a purely formal procedure. It is possibly the second claim that Searle finds especially provocative. As he says:

> It is an empirical question whether any given machine has causal powers equivalent to the brain. My argument against strong AI is that instantiating a program is not enough to guarantee that it has those causal powers. (p. 452)

But is this all? If so, then the supposed uniqueness of human cognition with respect to its separability from biological causation is not an important part of the strong AI thesis, so far as Searle's objections go. Yet a reading of Searle's subsequent publications, such as his review of Dennett and Hofstadter's book, *The Mind's I* (*New York Review of Books*, 29 April 1982), shows that he remains convinced of the low probability of successful duplication of the brain's relevant causal powers by nonbiological computational machinery. In the aforementioned review, he states:

> I believe that everything we have learned about human and animal biology suggests that what we call 'mental' phenomena are as much a part of our biological natural history as any other biological phenomena, ... Much of the implausibility of the strong AI thesis derives from its resolute opposition to biology; ... (p. 4)

The implicit argument in Searle's emphasis on biological causation runs as follows. If human cognition is the one biologically based activity claimed separable from its biochemical origins, then either there is strong evidence to support this claim, or the claim has a low probability by virtue of the large number of cases in which such separation is deemed impossible. But far from there being strong evidence to support this claim, there is actually an *a priori* argument against it. Roughly put, this argument premises that computation essentially involves nothing more than syntactic transformations, whereas human understanding involves the apprehension of meanings. But since strong AI attempts to create human cognition independently of a supporting biological system, using essentially nothing more than computation, it follows that no version of strong AI can be correct.

It is apparent that if Searle's *a priori* argument is successful his probabilistic argument is logically unnecessary. There are ostensibly good reasons for including it, however. First, in case the *a priori* argument fails some reason would remain to reject the strong AI thesis, barring the submission of strong evidence to support its alleged claim of uniqueness regarding human cognition. Second, the argument has rhetorical weight, setting a mood of suspicion against nonbiological alternatives to biologically based processes. And last, the argument can be offered as part of an explanation of those intuitive feelings some might harbour of the implausibility of the strong AI thesis.

These reasons carry only if the argument's assumptions are correct, however. A major assumption of the argument is that the strong AI thesis claims a special place for human cognition with regard to the possibility of nonbiological instantiation. But supporters of strong AI ought not be considered to be making any such claim. There are other types of specifically biological processes capable of nonbiological instantiation. Admittedly, cell division cannot be considered independent of cells, but what of others on

Searle's list of examples, such as photosynthesis, digestion, and lactation? It appears that any argument Searle might give here would need to be of the empirical (and debatable) sort that such processes have not been given a successful nonbiological instantiation. Missing from Searle's list of examples is the specifically biological process of respiration, and this omission is noteworthy, given the existence of heart-lung machines. Now Searle might argue that heart-lung machines are not completely independent of the biological organisms to which they attach during their actual operation, but at least the oxygenation of the blood (itself originally a biological process) occurs removed from the biological organism.

Of course Searle can claim that it is biological blood that is being oxygenated, so until a nonbiological substitute can be found to perform this function of biological blood, he still has an argument. It is hopefully apparent from the discussion, however, that he hasn't much of an argument. Taking the discussion further, Searle's inclusion of cognition in his list of examples, although necessary for his argument, appears somewhat strained. This may be due to considerations of space, however. That is, to state his point with reasonable brevity, Searle may have chosen not to make some sort of explicit distinction between, say, 'higher-level' biological processes, and the more often cited 'lower-level' examples. Given that the latter examples are more immediately identifiable as biological, Searle's brief list mentions them. But there are other 'higher level' processes that are associated with biological systems, for example domicile-building. To the best of our knowledge, only biological organisms engage in this activity. Yet it is within the realm of current technology that at least some domicile building can be done by automated, nonbiological electro-mechanical devices. It is beside the contended point that such devices are built by biological humans, since if a first-generation nonbiological, but genuinely cognitive, device is ever built, it is likely to be biological humans who will build it.

All of this shows that Searle has not much reason to be emphasizing biology in his attack on what he terms 'strong' AI. It may well be that only brains can produce thoughts, but he has given no reason to think this is so. What remains is his argument based on the assumption that computational machinery (in the usual sense of such machinery) is limited to the programmed task of syntactic transformations. If this argument fails, then he must find others if he is to maintain what he elsewhere terms his 'relentless exposure' of the 'preposterousness' of the strong AI thesis.

Miami University,
Oxford, Ohio 45056, U.S.A.

© ASHER SEIDEL 1988

EHRING ON PARFIT'S RELATION R

By MARK SIDERITS

DOUGLAS EHRING argues against Parfit's claim that Relation R (psychological continuity and/or connectedness) is what matters, that it violates Parfit's own non-triviality condition: since R is what matters, the difference between R's obtaining and not obtaining cannot depend on some trivial fact ('Survival and Trivial Facts', ANALYSIS 47.1, January 1987, pp. 50–4). I shall argue that Ehring's objection can be answered without abandoning any of Parfit's central claims. In doing so I shall also argue that Parfit is best understood as placing no causal constraints on R: what matters is Relation R regardless of how it comes about that R obtains.

Ehring considers three cases, all of which involve the use of a Scanning Replicator, a device that encodes on tape all relevant structural information concerning a brain and body, destroys that brain and body, and produces out of new matter a new body and a brain that is structurally isomorphous with the old brain. Because of this structural isomorphism, the person B who results bears Relation R to the original person A. (Parfit of course claims that A should regard this as about as good as ordinary survival.) In Case 1, the Scanning Replicator functions as intended. In Case 2, the tape which is made of A's brain is accidentally dropped on a second encoder, destroying this tape. But the fall causes the second encoder to malfunction in a highly improbable way – it produces a second tape which is an exact duplicate of the first. This tape is then used to produce person B, who bears R to A. Case 3 differs from Case 2 only in the following way: the fall of the original tape, while destroying it, does not cause the second encoder to produce a duplicate; instead an unrelated short circuit in this encoder causes it to produce a duplicate at just the moment that the first tape strikes it.

Ehring makes the following claims about these cases. If Parfit is right about what matters, then in both Case 1 and Case 2 the result is about as good as ordinary survival, but in Case 3 the result is about as bad as ordinary death. This is so because there is a causal constraint on Relation R: there must be some causal route from the psychological states of A to the psychological states of B. In (1) and (2) this condition is fulfilled, but in (3) it is not. The difference between Case 2 and Case 3 is, however, trivial. Hence Relation R does not, *pace* Parfit, meet the non-triviality condition.

In examining the psychological relations between A and B, Ehring confines his attention to the quasi-memory component of R. So it will be useful to focus on the causal condition that Parfit imposes on quasi-memory: the apparent memory is causally dependent, in the right kind of way, on the past experience (Derek

Parfit, *Reasons and Persons*, Oxford University Press, 1984, p. 220). A natural reading of this condition is in terms of the counterfactual,

> If A had not had experience *e*, then B would not have apparent memory *m*

(where it is open whether A and B are the same person or distinct persons). Now in Case 1 this counterfactual is fulfilled for each of B's apparent memories. For each such apparent memory is causally dependent on some structural feature of B's brain, each such structural feature is causally dependent on some tape state, each such tape state is causally dependent on some structural feature of A's brain, and each such feature of A's brain is causally dependent on some experience had by A. But in *neither* Case 2 *nor* Case 3 does this counterfactual seem to be fulfilled. In Case 2, in particular, it is highly implausible to suppose that this is true. Presumably in this case what causes the duplicate tape to have the tape states it in fact has, is the manner in which the first tape falls onto the second encoding machine. For the counterfactual to be fulfilled, Ehring must suppose that the first tape would not have fallen in just the way it did had A not had just those experiences A in fact had. And this is surely false. Under the circumstances envisaged, the falling of a *blank* tape might have caused the second encoder to produce a tape capable of duplicating all of A's brain states.

Thus it is not, as Ehring claims, the move from (2) to (3), but rather the move from (1) to (2) that eliminates the relevant causal dependence. It is thus open to Parfit to agree that the difference between (2) and (3) is trivial, yet also maintain that whether R obtains or not cannot depend on some trivial fact. He can do so by claiming that since the causal constraint on R is not satisfied in either (2) or (3), R does not obtain in either of these cases. And he can maintain that the difference between (1), where R does obtain, and (2) and (3), where R does not obtain, is deep. It is the difference between A's psychological states causing B's psychological states, and something else causing B's psychological states.

Consideration of cases like (2) and (3) may, however, lead one to wonder whether the causal constraint on R is justified. The facts in these cases are these: shortly after the Scanning Replicator destroys my brain and body, there will come into existence a new brain and body, and the resulting person will have the same degree of psychological connectedness with me now as normally obtains between successive days in the life of one person. Now we must suppose that I know these facts. Suppose in addition that because of an inoperable form of cancer, my present body would otherwise die in three months, but the new body will be missing this defect. Surely in such circumstances it would be rational for me to employ the Replicator in Case 1, where it serves as a reliable causal transmitter of Relation R. But if I know in advance that R will be preserved in Cases 2 and 3, why should I care that the resulting es will not be causally dependent on my past and iogical states? Of course we should ordinarily prefer n case be like Case 1 rather than Case 2 or Case 3. But d seem to be due solely to the fact that we cannot ordinarily on the sorts of accidents that, in (2) and (3), bring about the nded results. What we care about here is the result, not how the ult comes about. Parfit claims that what matters is Relation R ith any cause. These considerations suggest that we take him to mean that what matters is that there be some future person who is psychologically continuous and/or connected with the present person, no matter what the cause of the future person's psychological states. (It is only in discussing the quasi-memory component of R that Parfit introduces the more stringent causal requirement: that the apparent memory be caused by the experience.)

Ehring states that the resulting view of what matters is 'too liberal' (p. 54). He argues that we should then be required to view as about as good as ordinary survival the following situation: I am struck by a train and my body dies, but a moment later a person on some distant planet suffers a psychotic breakdown and as a result comes to be strongly psychologically connected with me just before I was struck by the train. Now I agree that we should not regard this as about as good as ordinary survival. But the problem here may not be the absence of causal connection between the two sets of psychological traits. Instead, the problem may stem from the fact that persons do not exist in a social vacuum.

Since this person lives on a distant planet, I assume that the shape and organic structure of its body will be quite different from my own. Consequently those of my acquired bodily skills in which I take pride will be, if not actually counterproductive, then at least quite useless to it. Likewise, the gender-class system of its society will no doubt be quite different from that of our culture. So those of my beliefs, desires, emotions, and emotional attachments that have been shaped by our culture's gender-class system will be inappropriate to it. We also cannot assume that anyone on this distant planet can speak English or read Sanskrit, so this person's newly acquired linguistic skills will prove quite useless for purposes of communication. Nor is it likely that this person will have at its disposal any of the resources needed to fulfil two of my major ambitions: to write a book on Indian semantics, and to restore a Victorian house.

Given all these and countless other facts concerning the mismatch between this person's newly acquired psychological traits and its social milieu, how should we expect its future psychological history to go? Those who have learned a foreign language through the method of total immersion in an alien culture will be able to appreciate the likely result. This person will shortly cease to care about, and so will lose, those beliefs, desires, values, ambitions, emotional attachments, skills and memories that I now care about.

That is, while psychological continuity might w
psychological connectedness would soon cease to h
hold only to an extremely diminished degree.

Now Relation R consists in both psychological contin
psychological connectedness. And as Parfit argues, psycholo
connectedness would seem to have at least as much value as psyc
logical continuity. I would not wish to survive as an anti-intellect
who demolishes Victorian houses and strives to bring about th
patriarchal theocracy depicted in Margaret Atwood's *The Handmaid's Tale*. Thus it is hardly surprising that I should view the situation envisaged by Ehring as about as bad as ordinary death. But suppose that what happens were this: a moment after my body dies, lightning strikes a tree in New Zealand, causing some of its molecules to be reassembled into a replica of my body and brain before the train accident. Psychological connectedness is nearly total: the resulting person lacks only my addiction to nicotine. Given certain desires that I now have, I might regard this as somewhat *better* than ordinary survival.

Parfit claims that what we care about in survival is both psychological continuity – that there be someone who has experience-memories of some of my experiences, whose character exhibits continuity with my own, and the like – and psychological connectedness – that, e.g. this person seeks to bring about my ambitions because he experiences them as his own. Neither component by itself seems to have terribly much value. But it does not follow that there is more to ordinary survival than these two components of R combined. R may well be what matters. And it may well be that R is what matters regardless of how it comes about that a future person is R-related to me.

Illinois State University,
Normal, Illinois, U.S.A.

© Mark Siderits 1988

WHY IS THERE SOMETHING RATHER THAN NOTHING?

By Noel Fleming

I

In his discussion of Locke's version of the cosmological argument, Leibniz remarks that from the premise that there has always been something it does not follow that some one thing has always been, i.e. that there is an eternal being (*New Essays*, 4, 10, 6). Anyone who sees this presumably also sees that from the premise that it is necessary for there to be something it does not follow that some one thing is necessary, i.e. that there is a necessary being; so Leibniz is not likely to make such a mistaken inference in the cosmological argument as he understands and variously states it. My aim here, however, is not to defend or criticize any of Leibniz's versions of the argument. What I want to do instead is to give a different version that derives from Leibniz and is meant to avoid the mistake of inferring the stronger necessity from the weaker or of confusing the one necessity with the other. It derives from Leibniz at least because it uses an unfashionable but I think not implausible premise taken directly from section 7 of the *Principles of Nature and of Grace*, the section in which Leibniz asks why there is something rather than nothing; and because this premise points to a Leibnizian presupposition of the argument that will be brought out below.

Here is the argument. Anything that is contingent could fail to exist, i.e. could not exist at all, could never exist; this is what it is to be contingent. So if everything is contingent then everything, i.e. each thing, could fail to exist; either as it were on its own, if its existence is independent of that of any other thing (as Leibniz, given his view of monads, would presumably take the existence of everything to be, if everything is contingent); or because of the non-existence of what its existence depends on, if some other thing or set of things is either necessary or sufficient for its existence, for any other thing could also fail to exist. And no two contingent things or sets of contingent things are such that, if nothing else existed and one of them failed to exist, then the other could not fail to exist; for the mere non-existence of something, independently of the existence of something else, does not necessitate the existence of anything. Whether or not there could be creation *ex nihilo*, or from nothing, creation *ab nihilo*, or by nothing, is on the face of it incoherent or impossible; and necessitation *ab nihilo* is no better. So if everything that could fail to exist did fail to exist then there would be nothing, if everything is contingent. So if everything is contingent then there could be nothing. But if there could be nothing then there would be nothing, for nothing is simpler and easier than something, i.e. it is simpler and easier, and certainly less is needed, for there to be nothing than for there to be something. (That nothing is simpler and easier than something is the premise from section 7 of the *Principles*.) But there is something, as each of us knows, if only in his or her own case. So there could not be nothing; so something is necessary, i.e. exists necessarily or is not contingent, for otherwise there could be nothing, and would be. In short: if everything were contingent then nothing would exist, but something does exist, so at least one thing is necessary.

So the answer to Leibniz's question that is given by this argument is that there is something rather than nothing because nothing is impossible, i.e. it is impossible for there to be nothing; and nothing is impossible because something is necessary. But in the order of the argument, in the order of knowledge, it is the other way around: we are supposed to know that something is necessary because we know that nothing is impossible, for we know that there is something, and we are meant to realize in the course of the argument that if nothing were possible, as it would be if everything were contingent, then there would not be anything at all.

II

We can bring out the presupposition of the argument that I referred to, and also see one natural or at least Humean way of rejecting the argument, by noting that from the premises that nothing is possible, and that nothing is simpler and easier than something, it does not follow that nothing rather than something exists. All that follows is that the simpler and easier of these two mutually exclusive but jointly exhaustive states of affairs is possible. So it is being taken for granted and could be made explicit in the argument that if the simpler and easier, which is nothing, can exist then it will exist, instead of the less simple and harder, which is something. Or, equivalently, it is being assumed that if the simpler and easier state of affairs in which nothing exists can obtain then it will obtain, instead of the less simple and harder state of affairs in which there is something. But why take this for granted, or take it to be true? The answer, at least in a Leibnizian context, is surely that if nothing is possible, and simpler and easier than something, then the existence of something, of anything or any things, is inexplicable; for if two mutually exclusive and jointly exhaustive states of affairs are possible, and one of them is simpler and easier than the other, and nothing else bears on the matter, then there is sufficient reason for the simpler and easier state of affairs to obtain, and no reason for the other to obtain. And what there is no reason for, or even what there is no sufficient reason for, is not so, for there is sufficient reason for everything's being so that is so.

Now none of this should bother Leibniz, as far as I can see. The argument with its presupposition is in a way just a different expansion from the one that he gives of a remark that he makes in *A Specimen of Discoveries About Marvellous Secrets of a General Nature*: 'If there were no necessary being, there would be no contingent being; for a reason must be given why contingent things should exist rather than not exist' (*Leibniz's Philosophical Writings*, ed. by G. H. R. Parkinson, Rowan and Littlefield, p. 76). Again, in section 7 of the *Principles*, he first lays it down that as we proceed from physics to metaphysics we must make use of the great principle of sufficient reason; he then says that the first question to ask once we have the principle is why there is something rather than nothing, for nothing is simpler and easier than something; as if, without the principle, we had better not even ask the question. But given that the argument presupposes the principle, we can reject the argument by not accepting the principle, and in the present context the principle is both doubtful in itself and obscure in what it requires.

We have no guarantee that there is an explanation for whatever happens, let alone an explanation for anything's happening at all. The existence of something rather than nothing may be inexplicable; what the argument shows is that it is inexplicable unless something is necessary. There is no sufficient reason for the existence of anything if everything is contingent; in a world of such pervasive contingency the existence of everything is, as we might put it, ultimately inexplicable, however ordinarily explicable any particular thing's existence is by reference to the existence (and nature and behaviour) of something else. But if anything is necessary then its existence is immediately ultimately explicable, for it could not have failed to exist; there is, trivially, as it were, a sufficient reason for the existence of a necessary being. So if we knew that something existed necessarily, and knew that it had given rise to or brought about the existence of contingent things, then we would have an ultimate explanation of their existence too; and the principle of sufficient reason tells us in effect that there is such an explanation. But it is not only that the principle may be false; what is required in the present case for the principle to be true, i.e. the state of affairs that has to obtain for there to be a sufficient reason for the existence of something rather than nothing, is opaque. Granted that the existence of everything is ultimately inexplicable if there is no necessary being, we nevertheless seem to understand why many contingent things exist – in that their existence depends on other contingent things, whose existence depends on still other contingent things, and so on – a good deal better than we understand even the possibility of the necessary existence of anything, or at least of anything that could explain the contingent existence of spatio-temporal things like us or the sun and the other stars. We have some grasp of necessity in this connection if only because we can contrast it with contingency; we seem to understand contingency well enough, and indeed have a lot of experience of it; and something is necessary if it is not contingent. But how could anything of the right kind be necessary? Hume's point that whatever we conceive of as existing we can also conceive of as not existing applies to Plato's demiurge if not to Plato's forms; and it is something relevantly like the demiurge, with efficient causal power, that would have to exist necessarily for the contingent existence of things like us and the stars to be ultimately explicable.[1]

University of California, Santa Barbara,
California 93106, U.S.A.

© NOEL FLEMING 1988

---

[1] I am very grateful to Francis Dauer for his comments on an earlier version of this paper.

EXPLICABILITY AND THE UNPREVENTABLE

By N. M. L. NATHAN

IN his *Philosophical Theology* (Indianapolis: Hackett, 2nd edn, 1980) James F. Ross develops an argument for the existence of an unpreventable and uncausable being. I will call it Argument (A). If sound, (A) proves that there is a being with two traditional attributes of God; as a natural theologian, Ross hopes that nothing in (A) will also force us to conclude either that this being has other attributes which God could not possess, or that it lacks attributes which God would have to possess. Like standard cosmological arguments and arguments from design, (A) assumes some general principles about explanation. But it differs from these arguments in that the principles about explanation which it assumes do not entail that there actually is a cause or explanation for anything but only that it is possible that certain things have explanations. Ross rejects pancausalism and the principle of sufficient reason and approvingly quotes Duns Scotus: 'I prefer to submit conclusions and premisses about the possible, for if those about the actual are conceded, those about the possible are conceded, but not conversely ...' (*De Primo Principio*, Ch. III; quoted *PT* p. 181, n. 34). I will try to show, however, that even if Ross's rather plausible principles about explanation are true, (A) runs into difficulties. And I will try to show that although a modified version of (A) would establish the existence of an unpreventable being, Ross's principles about explanation would themselves oblige us to deny that this being is uncausable, and thus deprive even the existence of an unpreventable being of any theological significance.

(A) runs as follows (I abbreviate Ross's own version, given on *PT* p. 174).

(1) If a proposition is false, then it is possible that there is an explanation of its falsity.
(2) All explanations of a proposition's falsity are either internal (in terms of the inconsistency of the proposition) or external (in terms of something's preventing the proposition from being true).
(3) 'There is an unpreventable and uncausable being' is consistent.
(4) It is impossible that anything should prevent the existence of an unpreventable and uncausable being.
(5) If 'There is an unpreventable and uncausable being' is false, it is impossible for its falsity to have an internal explanation. [From (3)]
(6) If 'There is an unpreventable and uncausable being' is false, it is impossible for its falsity to have an external explanation. [From (4)]

(7) If 'There is an unpreventable and uncausable being' is false, it is impossible for its falsity to have an explanation. [From (2), (5) and (6)]
(8) There is an unpreventable and uncausable being. [From (1) and (7)]

A natural objection is that (4) is false as it stands. Nothing can prevent the existence of an unpreventable being, but why should not something prevent an unpreventable being from also being uncausable? To meet this, we replace 'unpreventable and uncausable', in (4), and throughout (A), by 'unpreventable and unpreventably uncausable' and call the resulting argument (B). But even (B) is open to a radical objection. Consider this parallel argument, which I will call (C).

(1) If a proposition is false, then it is possible that there is an explanation of its falsity.
(2) All explanations of a proposition's falsity are either internal (in terms of the inconsistency of the proposition) or external (in terms of something's preventing the proposition from being true).
(3c) 'There is no unpreventable and unpreventably uncausable being' is consistent.
(4c) It is impossible that anything should cause the existence of an unpreventable and unpreventably uncausable being.
(5c) If 'There is no unpreventable and unpreventably uncausable being' is false, it is impossible for its falsity to have an internal explanation. [From (3c)]
(6c) If 'There is no unpreventable and unpreventably uncausable being' is false, it is impossible for its falsity to have an external explanation. [From (4c)]
(7c) If 'There is no unpreventable and unpreventably uncausable being' is false, it is impossible for its falsity to have an explanation. [From (2), (5c) and (6c)]
(8c) There is no unpreventable and unpreventably uncausable being. [From (1) and (7c)]

Since (B) and (C) have contradictory conclusions, we cannot accept that (B) is sound without denying the soundness of (C). But it seems impossible to find good grounds for preferring either one of these two arguments to the other. If we accept that one of the arguments is sound then we accept all its premisses as well as its logical validity, and if we deny the soundness of the other then either we deny its logical validity or we deny at least one of its premisses. But both arguments are logically valid, so to prefer one argument we must accept all its premisses and reject at least one premiss of the other argument. Suppose we initially accept (B). Which premiss of (C) can we reject? We cannot reject either (1) or (2), since these are also premisses of (B). (4c) seems analytically true. This leaves only (3c). But (3c) is no less plausible than

> (3b) 'There is an unpreventable and unpreventably uncausable being' is consistent,

which is again a premiss of (B). In the same way, if we initially accept (C), the only premisses of (B) which are not also premisses of (C) itself are

> (4b) It is impossible that anything should prevent the existence of an unpreventable and unpreventably uncausable being,

which again seems analytically true, and (3b), which is no less plausible than (3c). The parallelism seems to deprive both arguments of any cogency.

Can anything be salvaged from (B) on the assumption that (3c) is no less plausible than (3b)? Consider the argument we get if, throughout (B), we replace 'unpreventable and unpreventably uncausable being' by 'unpreventable being'. I will call it (D). If sound, (D) shows that there is an unpreventable being. Obviously (D) is logically valid. Can it be neutralized, like (B), by a parallel argument for the contradictory conclusion? If we try to construct such a parallel, (E), by substituting 'unpreventable being' for 'unpreventable and unpreventably uncausable being' throughout (C), then (4c) would be replaced by

> (4e) It is impossible that anything should cause the existence of an unpreventable being.

But (4e) is false. So (D) seems to stand. And although its conclusion is less interesting, from the point of view of natural theology, than the conclusion of either (A) or (B), the natural theologian could at least insist that there is nothing in (D) which is actually inconsistent with the existence of a being which is uncausable as well as unpreventable.

And yet I think that anyone who accepts Ross's principles about explanation will in the end find it difficult to accept that any being is uncausable. If we accept

> (1) If a proposition is false, then it is possible that there is an explanation of its falsity

then surely we should also accept

> (1a) If a proposition is true, then it is possible that there is an explanation of its truth.

And if we accept

> (2) All explanations of a proposition's falsity are either internal (in terms of the inconsistency of the proposition) or external (in terms of something's preventing the proposition from being true)

then surely we should also accept

(2a) All explanations of a proposition's truth are either internal (in terms of the inconsistency of the proposition's negation) or external (in terms of something's causing the proposition to be true).

But then we have the following argument, which I call (F).

(1a) If a proposition is true, then it is possible that there is an explanation of its truth.
(2a) All explanations of a proposition's truth are either internal (in terms of the inconsistency of the proposition's negation) or external (in terms of something's causing the proposition to be true).
(3f) 'There is no uncausable being' is consistent.
(4f) It is impossible that anything should cause the existence of an uncausable being.
(5f) If 'There is an uncausable being' is true, it is impossible for its truth to have an internal explanation. [From (3f)]
(6f) If 'There is an uncausable being' is true, it is impossible for its truth to have an external explanation. [From (4f)]
(7f) If 'There is an uncausable being' is true, it is impossible for for its truth to have an explanation. [From (2a), (5f) and (6f)]
(8f) There is no uncausable being. [From (1a) and (7f)]

It is of course true that (F) is mirrored by an argument, (G), for the conclusion that there is an uncausable being: (3f) is replaced, in (G), by

(3g) 'There is an uncausable being' is consistent

and (4f) is replaced by

(4g) It is impossible that anything should cause it to be true that there is no uncausable being.

But (4g) is false. To cause it to be true that something does not exist is the same as to prevent its existence. There is nothing in the argument which excludes the possibility that if there were any uncausable beings at least one would be preventable.

If (F) survives the parallelism objection, can it be rejected on other grounds? Ross himself defends a Scotist argument for the existence of an uncausable being (see *PT* pp. 176–80), and if this argument were sound we would have to reject at least one premiss of (F). But in fact the Scotist argument also suggests an equally plausible parallel argument for the conclusion that there is not even an unpreventable being. With 'x is potential' defined as '"x exists" is logically possible but x is not actual', the Scotist argument can be put like this:

That there is at least one Uncausable Producer is logically possible.
Whatever is logically possible is either actual or potential.
Whatever is potential is causable.
That there is an Uncausable Producer is not causable.
Hence, that there is an Uncausable Producer is not potential.
Therefore, at least one Uncausable Producer is actual, that is, exists.

The difficulty is that we could equally well argue as follows:

That there is no unpreventable being is logically possible.
Whatever is logically possible is either actual or potential.
Whatever is potential is causable.
That there is no unpreventable being is not causable.
Hence, that there is no unpreventable being is not potential.
Therefore, that there is no unpreventable being is actual, that is, no unpreventable being exists.

If we use the Scotist argument to criticize (F), we can hardly continue to defend even (D). And of course we can continue to defend (D) without being committed to the Scotist principle that whatever is potential is causable. (D) assumes only that whatever is potential is preventable.

I conclude then that Ross's argument (A) breaks down, and that although (D), which is a modified version of (A), may still establish the existence of an unpreventable being, this will not help the natural theologian, because premisses common to (A) and (D) ultimately commit us to the conclusion that this unpreventable being is causable.

School of Economic and Social Studies,
University of East Anglia,
Norwich NR4 7TJ

© N. M. L. NATHAN 1988

MARY MOTHERSILL ON AESTHETIC PLEASURE

By WILLIAM CHARLTON

IN *Beauty Restored*, Ch. X, Mary Mothersill argues that pleasure in general, and aesthetic pleasure in particular, while not a bodily sensation, is 'an inner state or feeling discoverable by introspection', and that the connection between this feeling and what we take to give it to us is 'causal' and 'contingent' (p. 281). If successful, her argument refutes anyone who wishes to apply to aesthetic pleasure the 'adverbial' account of pleasure advocated by Gilbert Ryle and

Anthony Kenny.[1] I shall try to show that her argument is inconclusive, and I shall suggest that it rests on a couple of misunderstandings.

According to Ryle and Kenny, getting pleasure from an activity is often simply engaging in that activity in a certain way. Professor Mothersill's main arguments for her own view appear in attempts to counter arguments for this account. She makes four points (pp. 281-3).

(1) You can have the feeling you get when you enjoy listening to a quartet without listening to the quartet. To show this Professor Mothersill asks us to think of a man who knows little about music and enjoys listening to a very similar quartet. This does not prove that the pleasure of listening to a quartet of a certain type is separable from listening to a quartet of that type. To establish that she might have urged that one could have the feeling while not listening to a quartet at all but merely believing one is, say in a dream or under the influence of drugs. But she sensibly refrains from this move which, as she observes (p. 296), would take her out of the real world into First Meditation-Land.

(2) Too intense pleasure may distract you from the activity. I find it hard to judge whether this is true and whether, if it is, pleasure cannot be adverbial. Might not too much enthusiasm interfere with a delicate task, and if so, would it follow that being enthusiastic is not acting enthusiastically? This consideration cannot, I think, carry much weight.

(3) The pleasure you get from listening to a quartet may last longer than the listening. This consideration will seem to show that the pleasure we get from an activity is separate from the activity only if we confuse two kinds of pleasure, enjoying an activity and being in a pleasant mood. When the quartet is over we may continue in a pleasant mood, but we cannot go on enjoying listening.

(4) Sometimes you can see plainly that someone is doing something, but it is only by inference that you can tell if he is getting pleasure from it. Professor Mothersill herself concedes that this could be true even if the Ryle-Kenny analysis is correct: You can see that I am gardening, but you cannot tell straight off whether my gardening is willing or reluctant.

One of the main objections to saying that the connection between pleasure and what gives it to us is contingent and causal, is that if that were so, we could be mistaken about what it is that is giving us pleasure. Professor Mothersill points out that many causal judgements are as good as incorrigible. When I drop a glass I know straight off that it is because it hits the floor that it breaks. Hence she can say that the apparent infallibility of many of our pronouncements on what pleases us is no proof that they are causal. But she also argues that we *can* be mistaken about what is giving us pleasure, and if she is right here, the Ryle-Kenny analysis is in serious trouble. I think, however, that her argument fails.

---

[1] Gilbert Ryle, *Dilemmas*, Cambridge 1954, Ch. IV; Anthony Kenny, *Action, Emotion and Will*, London 1963, Ch. VI. The account is applied to aesthetic pleasure in, e.g. my 'Aesthetic Reasons', *British Journal of Aesthetics* 1983.

I agree with her that Kenny concedes too much when he says:

> A schoolboy at a dormitory feast imagines that he is enjoying eating cold bacon and burnt potatoes when he is really enjoying the illicitness of his escapade.

But I do not agree that the schoolboy is mistaken about the cause of his pleasure (p. 310). If we distinguish between the object of someone's pleasure and its cause (Professor Mothersill resists this distinction, but without, I shall contend in a moment, good reason); if we say that the object is the thing which pleases and the cause is the past event or circumstance which makes it possible for the object to please: then the boy is right that the bacon or the feast[2] is the object of his pleasure but perhaps unaware that the illicitness of the escapade is its cause.

Arguing more positively, Professor Mothersill claims there is no clear phenomenological difference between rejoicing that *p* and enjoying ϕing; and goes on to suggest (p. 305) that I might think I am enjoying listening to the quartet when really I am rejoicing that I am a better musician than one of the players, or even that I am sophisticated enough to enjoy the listening. Her point cannot be established in this way because whether or not rejoicing and enjoying are phenomenologically similar, she herself admits they are logically heterogeneous. To establish her point she would have to show, *either* that I can think I am enjoying listening when in fact I am not enjoying that activity at all but another activity instead, *or* that I can think I am rejoicing at my superiority to the viola player when in fact I am not rejoicing at that but at some other supposed fact instead.

Professor Mothersill gives the example of 'eating buttered popcorn at the movies while enthralled by *Beneath the Planet of the Apes*' (p. 311), and a lesser aesthetician might try to use this to show that I can really be mistaken about which of the two activities I am enjoying. She herself rightly declares that they 'offer the possibility of a blissful synthesis'. There are other possibilities: the pleasant taste of the popcorn might be a causal factor helping me enjoy the activity of following the story in the film; it might even be that the concert-goer's rejoicing at his superiority to the viola player contributes in a quasi-causal way to his enjoying the activity of listening.

Professor Mothersill has one other argument which deserves mention. She says (p. 295) that questions of the form 'Why did he enjoy ϕing?' 'Why did he rejoice that *p*?' arise only when we should have expected the person would be disappointed that p or bored by ϕing. In such cases (she reasons) we feel a doubt, but about what? Not, surely, about whether 'He rejoiced at the victory' really states the object of the man's pleasure. So it must be about whether it really states the cause. If (as Ryle and Kenny would agree) the 'why?' question does not mean 'Convince me that the victory really was the object of the rejoicing' it must mean 'Convince me the victory was the cause'. Professor Mothersill here overlooks a third possibility. The question might mean: 'Convince me it was *reasonable* for him to rejoice at the victory'. The doubt arises, on her own admission, because we thought he favoured the side that was vanquished; if he did, his rejoicing seems unreasonable. This point about rationality is one to which I shall return.

---

[2] In some contexts we might distinguish between activities we enjoy and objects, events etc., which by pleasing us make the activities enjoyable; mistakes about either would be awkward for the Ryle-Kenny analysis, and the distinction is not important here.

I wish to conclude by suggesting that Professor Mothersill's opposition to the Ryle-Kenny analysis is based on a couple of misapprehensions. In the first place, she thinks this analysis employs an unsound or unhelpful concept of an *object* of pleasure. The 'object of pleasure for x at t' is wrongly taken to be some obscure kind of intentional object, when in fact it should be defined as 'what x takes to be the cause of his pleasure at t' (p. 300). I agree that the object of a man's pleasure is what he would say is giving him pleasure, and this is normally something public like a performance of a symphony or a mountain-ringed lake; at worst it is something like an incident in a book or an aesthetic property (in F. Sibley's, not in her sense of that term) of a vase or picture. But that does not stop it from being an intentional object, at least in the way in which (p. 284) the object of an emotion like anger is an intentional object. The object of my anger is the apparent slight, as Aristotle calls it, which I should give as that which is angering me, and this is normally a public episode. Moreover Professor Mothersill's use of the word 'cause' suggests that she thinks if I say 'It is the symphony (the personality of Lady Bracknell) that is giving me pleasure' I imply that the relation between my pleasure and what gives it to me is causal. That is absurd. Her theory of pleasure is not built into our ordinary aesthetic discourse any more than a causal theory about emotions is built into remarks like 'It is the behaviour of the Government that is making me angry'.

The second misapprehension is rather different. Professor Mothersill wants to define beauty in terms of pleasure, and she thinks (p. 279) that a Ryle-Kenny analysis of aesthetic pleasure would make that impossible. Far from making it impossible, it enables us to offer a definition which seems to me more plausible. For Professor Mothersill a thing is beautiful if it causes feelings of pleasure by means of certain of its features (those, she says, which are such that if anything else has them we shall not be able to tell it apart from that thing). How these properties are able to cause a non-bodily introspectible feeling remains a mystery to be solved, perhaps, by experimental psychology. I should prefer to say that a thing is beautiful if it is reasonable to enjoy activities in which one is aware of it (more precisely, if it has features which constitute aesthetic reasons for enjoying such activities). If Professor Mothersill's definition implies that aesthetic preferences are causally explainable, mine implies that they permit requests for rational justification, and that, I claim, gives it a considerable advantage.

The University,
Newcastle upon Tyne NE1 7RU

© WILLIAM CHARLTON 1988

CAUSATION, FORESIGHT AND COLLECTIVE RESPONSIBILITY

By GREGORY MELLEMA

SINCE the publication of Feinberg's acclaimed article, 'Collective Responsibility' (*Journal of Philosophy*, LXV, 1968, pp. 674-88), collective responsibility has become the object of intense scrutiny by a variety of writers, including Peter French, Virginia Held, David Cooper, Martin Benjamin, Stanley Bates, and Steven Sverdlik. In this essay I will identify and examine three theses about collective responsibility which seem frequently to be assumed or presupposed, but seldom explicitly stated, in philosophical discussions of collective responsibility.[1] While the first thesis places constraints upon what counts as collective responsibility in a way which is plausible and defensible, I will argue that the constraints placed by theses two and three are unreasonably limiting.

There is no perfect consensus in the literature as to how the concept of 'collective responsibility' is to be characterized. But for many writers the basic idea, at least, is that people can become responsible for states of affairs which they collectively bring about, either through action or inaction, in such a way that the responsibility for the state of affairs comes to be apportioned or accrued to the members of the collective.[2] Suppose, then, that each of several agents performs an act, call it a 'contributing action', by virtue of which he or she comes to bear responsibility for a particular outcome which they collectively bring about (where the contributing action can itself be the omission of an action). The three theses under consideration can then be stated as follows:

---

[1] An endorsement of all three can be found in Steven Sverdlik's recent article, 'Collective Responsibility', *Philosophical Studies*, LI, 1987, pp. 61-76 (see especially pp. 65-7).

[2] The term 'collective responsibility' is occasionally used to refer to a type of corporate responsibility which *cannot* be apportioned to the various moral agents. Regarding this type of responsibility I shall have nothing to say. Elsewhere, to avoid confusion, I have used the term 'shared responsibility' to mean what I here characterize as collective responsibility ('On Being Fully Responsible', *American Philosophical Quarterly*, XXI, 1984, pp. 189-93; 'Shared Responsibility and Ethical Dilutionism', *Australasian Journal of Philosophy*, LXIII, 1985, pp. 177-87; *Individuals, Groups, and Shared Moral Responsibility*, Peter Lang, Inc., forthcoming). Here, however, I shall simply employ the terminology of the majority of writers and hope that confusion is not thereby generated.

(T1): The agents are collectively responsible for the outcome only if each agent is responsible for performing the contributing act he or she performs

(T2): The agents are collectively responsible for the outcome only if the contributing act of each causes, or at least causally contributes to, the outcome

(T3): The agents are collectively responsible for the outcome only if each intends the occurrence of the outcome, or at least foresees the possibility of its being brought about.

Imagine a situation in which three youths deface a large outdoor mural by spraying it with paint. Each youth has a can of spray paint, each youth sprays paint at the mural, and the youths collectively bring about the defacing of the mural. What must be true if the youths are to be judged collectively responsible for this outcome? According to (T1), each youth must be responsible for performing his or her contributing action. According to (T2), the action of each youth must contribute causally to the outcome. And according to (T3), each youth must intend the occurrence of the outcome, or at least foresee its possibility.

(T1) is based upon the idea that an agent's responsibility for an outcome can be distinguished from the agent's responsibility for performing a contributing action. Surely it is reasonable to draw such a distinction. After all, one can be responsible for performing an act, such as attempted murder, when its intended outcome fails to occur (in which case one is certainly not responsible for the outcome). However, the main thrust of (T1) is that there can be no collective responsibility for an outcome unless each member of the collective bears responsibility for performing the contributing action in question. Thus, the three youths fail to be collectively responsible for defacing the mural if one or more of them fails to bear responsibility for spraying paint at it. For example, if one of the youths were forced at gunpoint to spray paint, it would no longer be true that the three are collectively responsible. Perhaps the other two would be collectively responsible for the outcome, but the three certainly would not.

It is hard to deny that responsibility for one's contributing action is a necessary ingredient for one's being a member of the collective responsible for the outcome. One who has not acted responsibly appears to be in no position to belong to such a collective. In this way the class of all the world's men appears to be a dubious candidate for qualifying as a collective responsible for the oppression of women. I conclude that (T1) is a plausible thesis about collective responsibility, and hence I find the Sartrean conception of collective responsibility (according to which, for example, all of a nation's people are collectively responsible for wars in which the nation is involved) somewhat extravagant.

(T2) asserts that there can be collective responsibility for an outcome only when the contributing action of each member of the collective causes, or causally contributes to, the outcome. (Feinberg appears to require, in addition, that it is the faulty feature of the action which contributes causally to the outcome for one to be considered at fault). By spraying paint at the mural each of the youths contributes causally to the defacing of the mural. In some circumstances the contributing action of each member constitutes a necessary condition for the occurrence of the outcome, as when three people are needed to lift a heavy object for the desired outcome to occur. But in order to contribute causally to an outcome, one's act need not constitute a necessary condition. The mural would have been defaced even if only one of the youths had sprayed paint at it, but from this it does not follow that not all of the youths have contributed causally to the outcome which in fact occurs.

No doubt in typical instances of collective responsibility the contributing actions of the collective's members contribute causally to the outcome. However, to insist that all instances of collective responsibility share this feature seems counter-intuitive. Suppose that during a prison riot a group of prisoners manages to seize control of the portion of the cell block which houses the electric chair. Having no weapons at their disposal, the decision is made to use the chair for purposes of executing a guard whom they have managed to capture. A sign on the wall instructs that the three switches, located in three different parts of the room, must be triggered simultaneously in order to deliver the fatal jolt. Therefore, three prisoners trigger the switches simultaneously and the guard is killed. By all appearances, then, these three prisoners perform acts which contribute causally to the death of the prison guard.

Nevertheless, the switches have recently been re-wired in such a way that one of the switches is no longer connected to the main circuit, a change designed to ease the trauma of the executioners (for the same reason that one member of a firing squad fires blanks). Thus, in the present scenario there is one prisoner whose acts play no causal role in the death of the prison guard. No one present knows that one of the switches has been disconnected, let alone which prisoner triggers it. Yet there is, in fact, one prisoner who does nothing to contribute causally to the outcome.

According to (T2), then, the collective responsibility for the death of the guard is attributed to the two prisoners triggering switches, as well as others who strap the guard into the chair, and so on, but not to the prisoner with the disconnected switch. Because of the fortuitous re-wiring of the switches, this prisoner turns out to escape being a member of the collective responsible for the guard's death.

(T2), in this manner, is based upon the notion that each of the members of the collective makes a causal contribution to the outcome. By all appearances the prisoner with the disconnected switch is a member of the collective which executes the guard. He desires that the guard be executed, he triggers the switch in an effort to bring this about, he has no reason to believe his switch is disconnected, and he cooperates with two fellow prisoners in what is intended to be a group endeavour. Why should the fact that his act does not contribute causally to the outcome exclude him from the collective which is responsible for it? It is hard to know on what grounds one might explain the rationale for this requirement, and in the context of this example it certainly has the appearance of being an arbitrary restriction.

One can, of course, contribute causally to an outcome in a manner different from the manner one intended. Suppose instead that all three switches had recently been disconnected. However, the very sight of the three prisoners triggering the switches induces in the condemned guard a massive cardiac arrest which proves to be fatal. In this scenario the three prisoners play a causal role in the guard's death by triggering their switches, but not in the way they intended. Thus, the prisoner with the disconnected switch in the last example plays a causal role after all, and by the lights of (T2) he is reinstated as a member of the collective (and this is the case even if Feinberg's refinements are employed). But, again, it is hard to see why the membership of the collective should be determined on the basis of fortuitous factors such as these. Perhaps it is plausible to restrict membership to those whose acts contribute to the outcome without insisting that they contribute causally to it, for there is a sense in which the prisoner with the disconnected switch contributes to the outcome (on this point see Zimmerman, Michael J., 'Sharing Responsibility', *American Philosophical Quarterly*, XXII, 1985, p. 120). But the present restriction, at any rate, seems unnaturally limiting.

One final restriction upon the scope of collective responsibility is stated in (T3). To be collectively responsible for an outcome, according to (T3), the occurrence of the outcome must either be intended or the possibility of its being brought about must be foreseen by each member of the collective (the latter clause insures that a group which collectively risks an outcome can be collectively responsible for its subsequent occurrence). Thus, if the youths are to be collectively responsible for defacing the mural, each must at least be aware of the possibility that the mural will be defaced as the result of their actions.

On the level of individual moral responsibility it is plausible to hold that a moral agent is responsible for an outcome only if he or she either intends it or is aware of the possibility of its being brought about. There are special cases, it is true, where it is through negligence that an agent is unaware of the possibility of its occurrence; in such cases it is surely possible for an individual to bear responsibility for it. In addition, there are cases in which it is reasonable to grant that an individual who stumbles blindly upon an unforeseen set of circumstances can bear responsibility for an outcome, even if the agent is not negligent of having failed to foresee it (a person might not be negligent for having failed to foresee that the telephone of the friend he is calling is being bugged by organized crime, but perhaps he can still be held responsible for thereby unknowingly conveying valuable information). It might not be evident to some that such cases are possible. Even if they are, however, one normally *qualifies* such ascriptions of responsibility with statements to the effect that the victim was at least partially the victim of circumstance. Accordingly, such agents are normally judged to be responsible for the outcome to a degree less than if the agent had foreseen the possibility of the outcome.

Aside from special cases such as these, then, it is plausible to hold that an individual is responsible for an outcome only if the individual either intends it or is aware of the possibility of its being brought about. But the situation is different with collective responsibility, for the members of a collective do not always share the same intentions concerning possible outcomes. Thus, an outcome can be collectively brought about by several individuals, some of which intend the outcome and some of which do not. Moreover, the members of a collective do not always foresee the same possible outcomes. Thus, an outcome can be collectively brought about by several individuals, some of which are able to foresee the possibility of its occurrence and some of which are not.

To see an example of this situation, suppose that Jones finds his next door neighbour's yard sculpture obnoxious and offensive. He purchases three cans of spray paint, and late one night he encourages his three teenage sons to spray paint at the sculpture. Arriving on the scene, the sons recognize at once that the old sculpture has just been replaced with another sculpture. They are uncertain what to do, for it occurs to them that the new sculpture (unlike the old) might be a costly, original work of art; nevertheless, they proceed to spray paint at it. In the end it is learned that they have indeed sprayed paint at a costly work of art.[3]

---

[3] It might be objected that this example fails to bring out the asymmetry between individual and collective responsibility. After all, if Jones had sprayed the costly work of art himself in the dark of night, believing it to be the original yard sculpture, he would arguably bear responsibility for defacing a costly work of art. Nevertheless, in this example Jones bears responsibility for the outcome (if he bears it at all) only in the qualified manner described above. He stumbles blindly upon an unforeseen set of circumstances, and he can plausibly lay claim to being at least partially the victim of circumstance. The same is not true in the original example, however, for the sons of Jones are able to recognize that the circumstances have been altered. Thus, neither they nor Jones is in a position to plead that they are victims of circumstance in a way which necessitates ascriptions of responsibility to them to be qualified in the same manner.

Who is collectively responsible for this outcome, the defacing of this costly work of art? Suppose it is assumed at the outset that Jones's three sons qualify for membership in this collective, and the question for consideration is whether Jones himself is likewise a member. According to (T3), of course, Jones does not qualify as an additional member of the collective; he neither intends the outcome nor foresees the possibility of its being brought about (nor can he plausibly be accused of negligence for his failure to foresee this possibility). He intends a different outcome, to be sure, but the outcome he intends does not occur. Thus, with respect to the outcome which does occur, Jones is excluded by (T3) from the collective responsible for it. He has no part in the collective responsibility which his sons bear for what happens.

No doubt this verdict is pleasing to Jones. By his sheer inability to foresee the possibility of the outcome's occurrence, he is able to escape membership in the collective responsible for it. But is this the correct verdict? Here I do not claim to have an absolutely convincing proof that Jones ought to be regarded as a member of this collective. My point, rather, is that it seems unreasonable to assume from the outset that he cannot possibly be a member. It is true that his sons foresee the possibility that a costly work of art will be defaced as a result of their actions, and he does not. But it is difficult to see why this fact, apart from all other considerations, should automatically allow him to escape membership in the collective responsible for the outcome.

Questions of collective responsibility are frequently complex and subtle, and there are important questions one would like to raise about Jones's role in the chain of events leading up to the actual outcome. (How much significance should be attached to the fact that the whole chain of events has its origin in Jones's desire to see his neighbour's sculpture painted? And what about the fact that the sons participated only because they were encouraged to do so by Jones? To what degree should this bear upon questions about his responsibility?). However, by endorsing (T3) one is committed to a prescribed verdict in a manner which makes such questions irrelevant.

Likewise, anyone involved in collectively bringing about a state of affairs slightly different from the one intended (where the possibility of the actual outcome is not foreseen) automatically escapes membership in the collective responsible for it, and all other questions about the person's involvement are rendered irrelevant, One need only establish that the actual outcome never occurred to one as a possibility (and that one could not have been expected to consider it), and (T3) serves to protect one from membership in the collective. Surely there are many individuals contributing to such outcomes (in times of civil turbulence, war, rioting, and so on) for whom (T3) provides precisely this type of relief. But is the result one really wants?

As a statement about collective responsibility, (T3) sounds initially appealing. However, upon closer inspection it commits one to judgments about the membership of collectives which seem very questionable. And, more generally, it forces one to decide questions of membership according to a process which hardly seems reasonable – automatically declaring those who neither intend nor foresee the possibility of the outcome's occurrence ineligible for membership. Perhaps (T3) can be suitably weakened, through the insertion of qualifying phrases, to accommodate cases where one desires individuals like Jones to be included, or at least eligible, for membership in the relevant collective. As matters stand, however, it appears that (T3), like (T2), places unreasonably limiting conditions upon what counts as an instance of collective responsibility. To the extent that discussions of collective responsibility tacitly appeal to either thesis, my suggestion is that one has reason to call into question the conclusions based upon them.[4]

Calvin College, Grand Rapids,
Michigan 49506, U.S.A.

© GREGORY MELLEMA 1988

---

[4] I am grateful to the editor for suggested improvements.

CLARK ON TRACING MENTAL IMAGES

By DAVID GORDON

IN 'Superman, The Image' (ANALYSIS 46.4, October 1986, pp. 222–4) Andy Clark argues that 'there may be less to mental images than meets the eye' (p. 224). His contention rests on a thought experiment in which one imagines Superman flying through the air.

> I can even, it seems, locate my image in the centre of the blank sheet of paper. It occurs to me to try to (as it were) trace the image. That is, I try to trace a picture of Superman by going around the edges of the projected image with ... [my] pen. But the project is a dismal failure. The puzzle then is how such a failure is possible. It need not be inevitable – the puzzle is genuine just so long as it is possible for even one person to be capable of visualizing the image yet incapable of tracing it' (p. 222).

Clark suggests several explanations, all of which, on the assumption that images involve some kind of operation closely akin to those involved in ordinary perception, seem to him implausible. He therefore suggests that this assumption about images is open to question.

Perhaps the assumption is questionable: but I do not think that, if it is, Clark's argument suffices to show this. The difficulty for the advocate of mental imagery arises because of the ease of the operation of tracing: if an actual picture of Superman were present on the paper, all one would have to do to *trace* it is to draw lines around its edges. If, in contrast, someone is asked to draw the outline of a picture of Superman, where the picture is not on the paper being used for the drawing, though the person has the picture before him, he has been presented with a task of substantially greater difficulty. The ability to draw accurately is, after all, one that many people lack; and the failure of someone to draw the outlines of a picture in these circumstances calls for no special explanation.

Clark's example assumes without argument that his case is one of essential similarity to physical tracing. (Although, in the passage cited, he speaks of his efforts to '(as it were) trace the image', he does not consider the implications of the bracketed phrase.) That is, if I form an image of Superman, and it seems to me that my image lies in the centre of a sheet of paper before me, it by no means follows that my image *is* in the centre of the sheet of paper. Notoriously, many defenders of mental images contend that images are not located in physical space at all. It may well seem to the imaginer of Superman that his image is on the paper; but, just as it is perfectly coherent for someone to claim that Macbeth's imagined dagger was not 'before him' in physical space but existed only 'in his mind', so might one contend that the image of Superman really is *not* on the paper. (Note that I have not argued that this view is correct but only that it is an alternative to Clark's view that the image is where it seems to the imaginer to be.)

If so, then the puzzle resulting from the failure to trace the outlines of Superman's images becomes merely a failure in accurate drawing. To the claim, 'But one *should* be able to trace the image, if one really has it; it's right there on the paper!', we can reply, 'Perhaps it isn't'.

Social Philosophy and Policy Center,
Bowling Green State University,
Bowling Green, Ohio 43403, U.S.A.

© DAVID GORDON 1988

TRACING SUPERMAN AGAIN: A REPLY TO CLARK

By SELMER BRINGSJORD

ANDY CLARK ('Superman, The Image', ANALYSIS 46.4, October 1986, pp. 222–4) has presented a thought-experiment which he thinks threatens the AI-relevant thesis that

> T The mode of representation underlying our image-forming ability involves something closely akin to the cognitive operation by which we perceive a real-world scene.

His thought-experiment is conducted if one does the following:

> E Sit down with a pen and a blank sheet of paper in front of you. Imagine Superman flying through the air. Locate this image in the centre of the blank sheet of paper. Then try to '(as it were) trace the image' – try to trace a picture of Superman by going around the edges of the projected image with the pen.

Most people who perform E fail miserably to trace the image, and Clark holds that this failure threatens T. I think Clark fails to show that E threatens T, because of the following counter-argument. Clearly,

> (+) That E threatens T is established by Clark only if the 'real-life' analogue E′ to E can be successfully performed.[1]

But the real-life analogue E′ to E *can't* be successfully performed. So by *modus tollens* Clark doesn't show that E threatens T.

At first glance this counter-argument seems obviously unsound, since tracing round a real object seems unproblematic: apparently one need only pick up a pen and use it to trace round the border of a picture, or round the border of an image projected onto a screen by a film projector[2] – a task, after all, which should be easy enough. But the catch is that the analogue E′ to E envisaged here is *not* an analogue. For notice that in E you are requested to do *more* than simply trace round the image: you are requested to trace round the image while looking (perhaps at times without wanting to) at the blank page and no doubt (peripherally) at a good many other things. This feature of E must be built into E′ if E′ is to be an analogue. (Surely two experiments E and E′ are analogous only if they share all relevant properties, and the property in question is certainly relevant, since it's a property which helps make E difficult to perform.) But then E′ might be something like

---

[1] That (+) is true, and why it is, was disclosed to me by an anonymous referee who made it clear that the crucial point in Clark's attack on T is that tracing round a real image has no counterpart for the image formed when conducting E.

[2] The example of the film projector is given by an anonymous referee.

E′ Sit down with a pen and a sheet of paper which contains a picture, but which is otherwise blank. Place your pen on the border of the picture and try to trace the picture *while at the same time* holding an image of Superman (or anything else) before your mind's eye. Notice that alternating between tracing and visualizing, however quick you do it, isn't requested: you must do these things simultaneously.

Most people fail miserably at this version of E′. Indeed (if the people I have tested are typical) most fail miserably at any such experiment (e.g. at tracing round the picture while *simultaneously* looking at something else, or carrying on a conversation, or doing mental arithmetic, etc.).

One might try to dodge my counter-argument by reformulating E so that the subject is no longer forced to do two things at once, and by taking as E′ scenarios wherein subjects can trace round a picture as they normally would if asked to do so. But when E is replaced by similar experiments which don't require that the subject look at the sheet when trying to trace, T appears to fare quite well. In fact, when E is replaced by such similar experiments, and then modified slightly so as to include the real-world analogue, T is apparently *supported*. For example, consider this replacement for E:

E″ Sit down at a desk with a pen in your hand and a blank sheet of paper on the desk. Place the tip of the pen on the paper, and then close your eyes and visualize Superman flying through the air (or move your eyes away from the paper and, while keeping them open, visualize Superman flying through the air). While visualizing Superman, without looking at the sheet, take your time and try your best to sketch him. Now do the same thing for a real object: get a clean sheet of paper and place, say, a bowl of fruit far enough away from your paper so that when looking at the bowl you can't see the paper. Now, while looking at the bowl, sketch it without looking at the paper.

When people conduct E″ they succeed: they produce two pictures. In fact, when E″ is modified so that what is visualized isn't Superman, but a bowl of fruit, or a plate of food, it's often the case that someone kept in the dark about the experiment is hard pressed to pick, after E″ has been conducted, the drawing based on the image (or the drawing based on the real object) from the two sheets of paper.

Let me conclude by saying that I have seen people whose success at E″ is only moderate, in the sense that which of their two pictures is which is evident: the one based on the image is not as detailed, etc. But the subjects achieving only moderate success can apparently be 'trained': if they practice by conducting E″ a few times, with different images and objects, their two sheets of paper get closer and closer in terms of detail, clarity, proportion, etc. This would seem to indicate that the difference in what counts as success in the two cases (the cases being the sketching of an image and the sketching of a scene that is actually before one) is a matter of convention. For when one evaluates the drawings produced by a 'trained' experimenter, one tends to apply, and to apply fairly, the same standards for detail, clarity, . . . to both drawings.[3]

Rensselaer Polytechnic Institute,
Troy, NY 12180, U.S.A.

© SELMER BRINGSJORD 1988

---

[3] I am indebted to an anonymous referee (referred to in the previous footnotes) who clarified the structure of Clark's argument, and to those who were willing to be subjects, especially in light of the fact that doing so in a sincere manner requires some rather strenuous concentration.

SUPERMAN AND THE DUCK/RABBIT: A REPLY TO GORDON AND BRINGSJORD

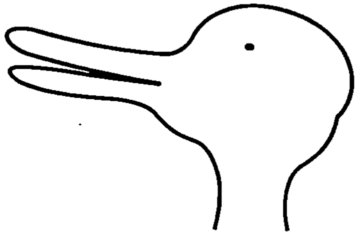
By ANDY CLARK

DOES mental imagery involve the same cognitive processes and operations as visual perception? In a previous paper (Clark [2]) I argued that it does not. In support of this claim, which goes counter to a fairly strong tradition in the literature on mental imagery (see e.g. Metzler and Shepherd [2], Kosslyn *et al.* [3]), I suggested a simple thought experiment designed to show that the operations we can perform on an image differ in functionally significant ways from those we can perform when actually perceiving a real object. The thought experiment, which I shall not repeat in detail here, involved comparing the ill-fated attempt to trace an *imagined* picture on a blank page, and the successful attempt to trace around the edges of a *real* picture. The picture used for the experiment was one of Superman flying through the air, hence the title of this piece.

Gordon and Bringsjord (above) both dispute the validity of the thought experiment. Both argue, in effect, that the *task demands* differ significantly in the case where the subject is required to 'trace' the image and that the difference in task demands invalidates the comparison between the two cases. Gordon says that 'Clark's example assumes without argument that his case is one of essential similarity to physical tracing' but insists that in reality '[the image-tracer] has been presented with a task of substantially greater difficulty'. The extra factors which Gordon fixes on are (a) general drawing abilities and (b) the location of the image in the centre of the page. Likewise Bringsjord notes that, in the image-tracing case one must not *simply* trace the image but trace around it while looking at the same time at the blank sheet of paper and perhaps at other peripheral things also. For this reason he too denies any valid comparison between the two cases.

In effect, the objection is that since there are extraneous factors likely to cause a difference in performance on the two tasks, we cannot conclude that perception and imaging do not share most of the same cognitive processes and operations. I do not believe that the factors which Gordon and Bringsjord point to cause significant and uncontrollable differences in the task demands. But rather than conduct a mere exchange of intuitions here, I propose instead to offer a new example drawn from the experimental psychological literature in which all such factors are controlled. The example (which was drawn to my attention by David Young of the Sussex University School of Cognitive Sciences) involves the operations we can and cannot perform on mental images of ambiguous pictures.

Chamber and Reisberg [4] conducted the following experiment. As a means of testing the hypothesis that visual imagery and visual perception are largely equivalent as regards underlying cognitive operations, they asked subjects to reverse a mental image of an ambiguous figure. The figure they chose was the duck/rabbit shown below. Most of us, faced with this picture and told to try to find an alternative interpretation to 'duck' can quickly discover the right-facing rabbit profile. This reversal, it seems, is not possible with the mental image.



The experimenters chose a group of subjects ranged across a scale of 'image vividness' as measured by Slee's Visual Elaboration scale (Slee [5]). The subjects, who did not already know the duck/rabbit picture, were trained on related cases (Necker cubes, face/vase pictues) to ensure that they were familiar with the phenomenon in question. They were then briefly shown the duck/rabbit and told to form a mental picture so that they could draw it later. They were then asked to attend to their mental image and to seek an alternative interpretation for it. Hints were given that they should try to shift their visual fixation from e.g. lower left to upper right. Finally, they were asked to *draw* their image and to seek an alternative interpretation of their own drawing. The results were surprising.

> Despite the inclusion of several 'high vividness' imagers, none of the 15 subjects tested was able to reconstrue the imaged stimulus. ... In sharp contrast, all 15 of the subjects were able to find the alternate construal in their own drawings. This makes clear that the subjects did have an adequate memory of the duck/rabbit figure and that they understood our reconstrual task ([4] p. 321).

Three further experiments were conducted to control for various factors, and all supported the same conclusion. Images, it seems, are *not* capable of ambiguity in the way a real picture is. When subjects *familiar* with the duck/rabbit claim to be able to perform the switch, we may assume (in the light of their inability to *discover* ambiguity thus) that they in fact *replace* one interpreted image by *another* interpreted image. In general, some of the kinds of operation we can perform on a real perceptual input are simply unavailable to be performed on an image because:

> Images are not quasi-pictorial in the sense of needing some interpretation. Instead, an images referent is specified by the imaginer. Thus although an image might in some fashion resemble more than one content ... this does not result in ambiguity because it is not by resemblance that images refer ([4] p. 325).

The clear conclusion is that imagery and perception, although they may share various properties associated with e.g. the use of representations preserving the metric properties of space (see Metzler and Shepherd [2]), do *not* share low level processing paths and operations. Chambers and Reisberg are explicit: 'Our results can thus be taken as arguing against the claim that imagery and perception share a common processing path' ([4] p. 326).

Returning to Gordon and Bringsjord's criticisms, I submit that in the newly cited example there are no extraneous factors (e.g. the need to attend also to interfering external stimuli – like the blank page and surroundings) making the operation on the image more difficult than its real-life analogue. The only significant difference in the new case is precisely that one task involves an image and the other a drawing of the image.

I concluded the previous paper by claiming that there was less to mental images than meets the eye. I am now willing to make a slight revision. There is *more* to mental images than meets the eye; for the image is bound to an interpretation in a way the percept is not. Apart from this minor shift, I hold myself to have been on the side of the angels all along!

School of Cognitive Sciences,
University of Sussex,
Brighton BN1 9QN

© Andy Clark 1988

REFERENCES

[1] A. Clark, 'Superman; The Image', *Analysis*, 46.4 (October 1986), 222-4.
[2] J. Metzler and R. N. Shepherd, 'Transformational Studies of the Internal Representation of Three-Dimensional Objects', in R. L. Solso (ed.), *Theories of Cognitive Psychology*, the Loyola Symposium (Lawrence Erlbaum, Hillsdale, N.J., 1974).
[3] S. M. Kosslyn, S. Pinker, G. Smith and S. P. Schwartz, 'On the Demystification of Mental Imagery', *Behavioural and Brain Sciences*, 2, 4 (1979), 535-48.
[4] D. Chambers and D. Reisberg, 'Can Mental Images be Ambiguous?', *Journal of Experimental Psychology: Human Perception and Performance*, II, 3 (1985), 317-28.
[5] J. Slee, 'Individual Differences in Visual Imagery Ability and the Retrieval of Visual Appearances', *Journal of Mental Imagery*, 4 (1980), 93-113.

NATHAN ON EVIDENTIAL INSATIABILITY

By HOWARD SIMMONS

IN a recent article, N. M. L. Nathan discusses the question of possible infinite regresses in the justification of our beliefs ('Evidential Insatiability', ANALYSIS 47.2, March 1987, pp. 110-15). He distinguishes three functions which the quest for justification may serve: 'The first function, which I call the satisfaction of local curiosity, can, I argue, be served by a justified belief which requires no regressive steps on the believer's part, and the second, or stabilizing function, can be served by a justified belief which requires the believer to make only a manageable number of regressive steps' (p. 111). But the third function, that of removing 'global suspicion', does, according to Nathan, necessitate an infinite regress which appears to make genuine justification impossible. I shall argue, however, that there is in fact no regress in the global suspicion case. The latter may be assimilated to the local uncertainty case, in that the same kind of justification is available for it, at least in principle.

Let us first look at the local uncertainty case. Here the function of justification is, says Nathan, that of making up one's mind about the truth-value of some particular proposition. Call this proposition *p*. Suppose the evidence for *p* is *q*. Might I not also want to seek evidence for *q* and also evidence for *that* evidence and so on *ad infinitum*? The answer, says Nathan, is obviously 'no', if my only reason for seeking evidence for *any* proposition is to make up my mind about its truth-value. In using *q* as evidence for *p*, I must be able to make up my mind that it is true. Thus, on the stated assumption, I do not need evidence for *q*, and so no regress is generated.

For the purposes of this article, we can ignore Nathan's 'belief stabilizing' function and move immediately now to his third function of justified belief – that of removing 'global suspicion'. In this case, I am concerned that my *entire* set of beliefs may not contain only truths. How can I assure myself that it does? Presumably by finding some proposition *q* which is evidence for this. Now Nathan presents us with the following dilemma: either I believed *q* before setting out on my justificatory enterprise or I did not. If I did, then *q* itself belongs to the very set of beliefs that I am trying to justify. But in that case, it cannot after all be used to support that set, on pain of a vicious circle. If, on the other hand, I did *not* believe *q* before embarking on my justificatory enterprise, I am presumably going to have to believe it now, if I want to cite it as evidence. But in that case, I have a new object of global suspicion – the original set of beliefs *plus q*. How can I be sure that *this* set contains only truths? Presumably by citing as evidence a new proposition q'. But then exactly the same reasoning can be applied in relation to q'. In this case, unlike in that of 'local uncertainty' or 'belief stabilization', we do seem compelled, says Nathan, to accept an infinite regress.

To this one might make the following objection. It is granted that I am able to make up my mind that *q*, the evidence used to support the original set of beliefs, is true. So just because it *does* support the original set, I am also able to make up my mind that all the members of this set are true. So why should there be any problem about believing the *conjunction* of *q* with the original set? In other words, it simply does not seem to be the case that the new set is a legitimate object of suspicion. (Nathan's own discussion of 'local uncertainty' makes it clear that he does not think that we are under any general *obligation* to seek evidence for what we have already made up our minds about.)

In response to this, Nathan concedes that if I *know* what proposition *q* I would use to support my original set of beliefs, then – since I must believe it in order to use it as evidence – I cannot wonder whether or not it is true. So if I am logical, I will also be sure in that case that the result of conjoining it to the original set is also true. But he points out that I may *not* know what would serve as a suitable *q*, in which case the doubt in question will arise, and with it the infinite regress. This then essentially completes Nathan's argument for maintaining that there is a genuine worry about infinite regresses in the global suspicion case.

But there seems to be something wrong here. Nathan maintains that there is a regress in the case where I do not know what *q* is – the case where I do not know what evidence I can use to support the thesis that all the members of my original belief-set are true. But it is in fact very difficult to see how the alleged regress can arise. If I do not even know what *q* is, how am I to go about looking for a proposition q' that will support the result of conjoining it to the original set? Surely no q' can really enter the picture, and thus no regress can get going. To make this clearer, let us return to the local uncertainty case. Here a regress was avoided by pointing out that, if I only need support for propositions about whose truth-value I have not made up my mind, then I will not need support for the proposition *q* which I am using to justify the original proposition *p*, since I will already have made up my mind that *q* is true. But one can now imagine someone arguing, in the manner of Nathan when discussing the global suspicion case, as follows: I may not *know* what proposition *q* I could use in support of *p*. I can therefore legitimately wonder whether or not it would be true. This would create the need for a further proposition *q*′ to justify it, and thus launch us on an infinite regress. Clearly this argument would be very unconvincing. For if I do not *know* what *q* is, then surely I cannot even go about looking for a suitable proposition *q*′ with which to support it. So no *q*′ ever enters the picture, and no regress ever gets underway. The situation is precisely the same in the global suspicion case. And it is important to stress this analogy with the local uncertainty case, since without it one might not be fully convinced that there is no regress in the case of global suspicion. There is just as much reason to accept a regress in the former case as in the latter. But Nathan himself denies – and very plausibly – that there is a regress in the former case. So he should also deny that there is one in the latter. This does not mean that there is no sceptical problem regarding the justification of one's entire belief-set. But the sceptical problem – if it exists – will be merely the very fact that I do not know what evidence to use in support of the belief-set, not the existence of a regress supposedly generated by this lack of knowledge.

In fact, when I say that this would be the sceptical problem, this is only roughly true. For my entire set of beliefs almost certainly contains some falsehoods, and therefore I should not *want* a proposition which would support all of them. But there may be some subset of them – say all those that satisfy a certain condition *C* – which I am inclined to give my assent to, and of whose truth, or at least warranted assertability, I would like to be assured. ('*C*' might describe a certain method of acquiring beliefs.) I might then look for a proposition *q*, of which I am certain, which would convince me that all beliefs satisfying condition *C* are true, or at least warrantedly assertable. This seems to be closer to an acceptable kind of global justification than what Nathan describes. And its pattern is essentially the same as that involved in the local uncertainty case. In both cases I use a proposition *q* of which I am already certain, to convince me of something else. In the local uncertainty case, the 'something else' is the truth of some other proposition. In the global case, it is the warranted assertability of a set of propositions.

White Acre, Pitton, Salisbury,
Wilts SP5 1DG

© HOWARD SIMMONS 1988

NATHAN ON GLOBAL JUSTIFICATION

By ROBERT P. MCARTHUR

N. M. L. NATHAN suggests that regressive concepts of justification are particularly unhelpful when applied to suspicions or doubts about the truth-value of one's entire system of beliefs ('Evidential Insatiability', ANALYSIS 47.2, March 1987, pp. 110–15). Suppose one wants to be assured that one's belief-system contains only true beliefs. To be so assured requires some evidence, and having such evidence in the form of proposition q requires one to actually believe q. Nathan rejects the possibility that such an evidential proposition q could already be a member of one's belief-system, because 'you cannot assure yourself that a conjunction is true by citing as evidence for its truth a proposition which is a member of the same conjunction' (p. 114). And if q isn't already believed, then adding it to one's beliefs produces a new set of beliefs about which similar suspicions can be formulated. Evidence for the new set will be evidence for q itself (since q counts as the evidence for the original set); the new evidential proposition can't, on Nathan's grounds, already be believed; and the problem begins anew, and so on, *ad infinitum*.

While Nathan subscribes to scepticism about justifying entire belief-systems regressively, he is more sanguine about the possibilities of justification for single propositions. He argues that the search for evidence q for belief p might take one through a regression of beliefs q_1, q_2, $q_3, \ldots$ which could *terminate* with belief q_n which itself does not require any further justification. 'It would do', he states, 'for you to believe that, quite independently of whether you have evidence for it, q_n is a proposition *which you cannot help believing*' (p. 112; emphasis added).

Consider, then, a proposition of the following sort:

(*e*) All my beliefs are true because they possess property P.

Suppose, for some P, that I believe (*e*) and, hence, that (*e*) is already part of my entire system of belief. Not only does (*e*) provide evidence for the truth of all of my other beliefs, since I believe (*e*), it provides evidence for that belief (i.e. itself) as well. Hence given that a proposition of the form (*e*) is a member of my system of beliefs, I could have the desired global assurance that all the beliefs in the system are true.

But I may well want some additional assurance that the propositions on which (*e*) is based are justified, that is, I may also want justification for both of the following:

(1) All my beliefs possess property P,

and

(2) Any proposition possessing property P is true.

Assuming that P is reasonably well-defined, (1) could be dealt with by simple inspection of my beliefs, one by one. The justification of (2) may seem to threaten to send us down the regressive spiral so favoured by the sceptic. But on Nathan's account it would do to find that (2) is a proposition which I believe I cannot help believing (or is based on one which I cannot help believing). And if that were the case, I would still be within my original belief-system and no further regression would be necessary for complete justification.

And what might this property P be? It is easy to imagine a person believing that all of his beliefs are divinely inspired, or are in conformity with the teachings of Big Brother, or are similarly blessed with truth in some (to him) irresistible fashion. Such a person would, on Nathan's theory, be justified in feeling that all of his beliefs are justified. But surely he would be, in fact, still far from being justified in this feeling.[1]

Colby College,
Waterville, Maine 04901, U.S.A.

© ROBERT P. MCARTHUR 1988

---

[1] I am grateful to the editor for his helpful comments on an earlier version of this note.

WILLIAMS ON WHAT THE PRESIDENT KNEW

By THOMAS P. FLINT

IN reviewing C. J. F. Williams's *What is Existence?* (Oxford: Clarendon Press, 1981) in *The Philosophical Review* (93 (1984), pp. 131–4), I recently suggested that the type of analysis Williams offers of sentences such as

(4) President Reagan knows that David Pears exists

is susceptible to counterexamples. According to Williams, (4) can be paraphrased by

(5) For some θ, both David Pears alone θ's and President Reagan knows that for just one x, θx.

The counterexample which I offered was directed explicitly not at (5) as an analysis of (4), but at

(2) For some θ, both David Pears alone θ's and President Reagan believes that for just one x, θx

as an analysis of

(1) President Reagan believes that David Pears exists,

but was intended as an objection to either analysis. In his recent response ('Knowledge, Belief and Existence', ANALYSIS 47.2, March 1987, pp. 103–10), Williams provides the following lucid and accurate summary of my counterexample:

> The circumstances envisaged are those in which David Pears, and no one else, is in the White House and no one at all is in the Capitol, whilst President Reagan believes, mistakenly, that Howard Baker alone is in the Capitol and that no one at all is in the White House. Thus,
>
> (3) David Pears alone is-either-in-the-White-House-or-in-the-Capitol and President Reagan believes that for just one x, x alone is-either-in-the-White-House-or-in-the-Capitol
>
> is true, and since (2) is an existential generalization of (3), (2) is true also. But in these circumstances there is no guarantee that (1) is true (p. 103).

Williams agrees that this counterexample shows (1) and (2) to be non-equivalent, and suspects that I would try to argue '*a pari*' that it also shows (4) and (5) to be non-equivalent. To this, Williams replies, '*nego paritatem*'. Reminding the reader that his book nowhere claims that (1) and (2) are equivalent (an interpretation, he says on p. 104, 'which Professor Flint wishes to foist on me'), he concedes that he is committed to the equivalence of (4) and (5), but fails to see how my counterexample calls that equivalence into question:

> In the scenario suggested by Flint, ... (5) is no more likely to be true than (4). Since the President's belief about Howard Baker is, *ex hypothesi*, mistaken, we have good Gettier-type reasons for denying that he knows that for just one x, x is-either-in-the-White-House-or-in-the-Capitol. So Flint has done nothing to disprove the material equivalence, at least, of (4) and (5) ... (p. 104).

In response, let me first plead guilty (or at least *nolo contendere*) to the charge of foisting. Williams' discussion was indeed of knowing, not believing.

However, on the more important point, it seems to me that no apologies or retractions are in order, for (*pace* Williams) it seems evident to me that my counterexample *does* disprove the thesis that (4) and (5) are equivalent. Williams purports to deal with this counterexample by alluding to the Gettier problem; once the hallowed name of Gettier is invoked, he implies, the efficacy of my example will have been exorcised, for we will no longer be tempted to think that (5) is true. Williams does not, alas, grace us with an explanation as to precisely how or why this incantation is supposed to have this cleansing effect, but I think it is fairly easy to divine what he has in mind. (5) will no longer beguile us because the second conjunct of its relevant instantiation — viz.,

(6) President Reagan knows that for just one x, x is-either-in-the-White-House-or-in-the-Capitol —

will no longer strike us as plausible. Why? Because, Williams implies, the Presidential belief upon which (6) depends lacks the kind of justification required for knowledge, for it is based upon the President's mistaken belief about Howard Baker.

It is to this last point that I must object. True, the President's belief that

> (7) For just one *x*, *x* is-either-in-the-White-House-or-in-the-Capitol

could be based upon his mistaken belief about Howard Baker, and hence lack the proper justification, but the situation I envisaged hardly *requires* that it be so based! It could easily be that the President's acceptance of (7) is totally independent of his confusions concerning Baker.

To see that this is so, let us augment the suggested scenario in the following way. Suppose that, in a fit of Nixonian paranoia, Reagan has installed an elaborate alarm system which informs him, wherever he happens to be, as to the total number of people in the White House and the Capitol. When one light goes on, the President knows, exactly one person is either in the White House or in the Capitol; two lights, two persons; and so on. Suppose that, having installed this system and tested it rigourously, President Reagan orders that the Capitol and the White House be evacuated. Within a few minutes, his alarm system tells him that his order has been carried out. Suppose that Reagan now sends Howard Baker to the Capitol. Five minutes later, one light on the President's alarm system goes on. Seeing this, the President forms the belief that (7), and also forms the belief that Baker has arrived at the Capitol. In this latter belief, though, he is mistaken; in fact, the light has gone on because the crafty David Pears has surreptitiously entered the White House.

In this situation, the President does indeed have a mistaken belief about Howard Baker. But it seems clear that this mistake by no means renders the President's belief that (7) unjustified. Given his prior experience of the reliability of his alarm system, the President is fully warranted in believing that (7). But if so, it would seem that (6) is true. And if (6) is true, then so is (5). But surely it would be consistent with our scenario that Reagan not know that David Pears exists. In other words, (4) could easily be false in the situation imagined. So our counterexample does indeed suffice to show that (4) and (5) are not equivalent, *quod erat demonstrandum*.

University of Notre Dame,
Notre Dame, Indiana 46556, U.S.A.

© THOMAS P. FLINT 1988

HOW MUCH DID THE PRESIDENT KNOW?

By C. J. F. Williams

THE point at issue between Professor Flint and myself now boils down to this. In the new circumstances he envisages (see pp. 61–3 above) I grant him that it is true that

> David Pears and no one else is-either-in-the-White-House-or-in-the-Capitol

and that

> President Reagan knows that, for just one x, x is-either-in-the-White-House-or-in-the-Capitol.

From these two premisses it follows that

> (5) For some θ, both David Pears alone θ's and President Reagan knows that, for just one x, θx.

According to my analysis, (5) is equivalent to

> (4) President Reagan knows that David Pears exists.

However, according to Flint, whereas (5), in the circumstances envisaged, is true, (4) is false.

It is just this that I wish to dispute. Both (5) and (4) are, in my view, true. There is, after all, just one man who is-either-in-the-White-House-or-in-the-Capitol; and the President knows this. To know that a description of this sort is uniquely satisfied, is to have the means of referring to the object which satisfies the description. The President can say things about 'the man who alone is-either-in-the-White-House-or-in-the-Capitol'. He could even, in other circumstances, give him a name. He knows, therefore, that this man exists. He does not, of course, know that the man in question is David Pears — because he has never heard of David Pears. Moreover, he thinks, mistakenly, that the man in question is Howard Baker. But I had already argued on pp. 97, sq., of *What is Existence?* that a President need not know a man's name in order to know that that man exists. And I went on to argue that it is no objection to saying that a President knows that a man exists if he confuses him with someone else. President Carter, I supposed, knows of the existence of the author of the fortieth chapter of the book we call the Book of Isaiah, although he confuses him with a prophet called 'Isaiah' who was alive in the year that King Uzziah died. The case for President Reagan's knowledge of the existence of David Pears, in the circumstances envisaged, can be argued *a pari*. I do not need to repeat the argument. I look forward to another *confiteor* from Professor Flint.

University of Bristol,
9 Woodland Road,
Bristol BS8 1TB

© C. J. F. Williams 1988

The Journal of Philosophy

SUBSCRIPTIONS

Individuals	$25.00
Libraries and institutions	$40.00
Students, retired/unemployed philosophers	$20.00
Postage outside U.S.	$ 6.00

BACK VOLUMES

Published bi-monthly 1904–1976, monthly thereafter.

All back volumes and individual issues are available, also a Cumulative Fifty-year Index, 1904–1953 and a Ten-year Supplemental Index, 1954–1963. Back volumes 1976–present also available in microform.

Please inquire for price lists and shipping charges on back orders.

709 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

METAPHILOSOPHY

Edited by Terrell Ward Bynum

Metaphilosophy publishes articles and book reviews in a wide range of philosophical topics including the foundations, scope and function of philosophy, aspects of philosophical systems, applied philosophy, the sociology of philosophy, the teaching of philosophy, presuppositions of philosophical schools, and the relation of philosophy to other disciplines.

January, April, July, October
(two single issues and one double issue per year)
Subscription Rates Volume 18, 1987
Individuals: £20.50 (UK) £25.65 (overseas) US$43.25 (N. America & Japan)
Institutions: £38.50 (UK) £48.00 (overseas) US$80.00 (N. America & Japan)
Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: Journals Department
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Basil Blackwell · Oxford and New York

THEORIA

A Swedish journal of philosophy
Edited by Bengt Hansson, Ingmar Persson and Krister Segerberg

Volume 52, parts 1–2

Ingvar Johansson Levels of intension and theories of reference **Michael J. Wreen** The Power of Potentiality **Michael H. Robins** Intention and control **Georg N. Schlesinger** Relevance **Gilead Bar-elli** and **David Heyd** Can revenge be just or otherwise justified? **Sven Ove Hansson** Individuals and collective actions **John Bishop** Review of Sajama, S: Idea, judgement and will **Douglas N. Walton** Review of Talja, J: Studies in possibility, ability and not-doing **José A. Lopez Cerezo** Review of Tuomela, R: A theory of social action.

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back volumes should be sent to *Editor,* THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 100:– or US$ 20.–, for three years SwCr 240:– or US$ 46.–

Teaching Philosophy

Tenth Anniversary Special

Teaching Philosophy, edited by Arnold Wilson of the University of Cincinnati, is celebrating ten years of publication by offering subscriptions to new individual subscribers at a special anniversary rate of only $10 per year. The regular subscription rate is $17 for philosophers and $38 for institutions. Orders must be pre-paid and received by November 30, 1987 to get this special low rate.

* * *

To help you discover the latest trends in the teaching and learning of philosophy, ***Teaching Philosophy*** publishes articles, discussions, reports, and reviews on these and related topics:

- theoretical issues in the teaching of philosophy
- experimental and interdisciplinary philosophy courses
- evaluation of teaching and assessment of learning in philosophy
- new books and audio-visual materials

* * *

Send orders to: *PHILOSOPHY DOCUMENTATION CENTER, Bowling Green State University, Bowling Green, OH 43403-0189.* ☎ *(419) 372-2419*

Law and Philosophy

An International Journal for Jurisprudence and Legal Philosophy

Editor

Alan Mabe, *Dept. of Philosophy, Florida State University, Tallahassee, USA*

Law and Philosophy is a forum for work in law and philosophy which is of common interest to members of the two disciplines of Jurisprudence and Legal Philospphy. It is open to all approaches in both fields and to work in any of the major legal traditions – common law, civil law, or the socialist tradition. The editors encourage papers which exhibit philosophical reflection on the law informed by a knowledge of the law, and legal analysis informed by philosophical methods and principles. Among others, the following topics are dealt with: theories of law, justice and law, morality and the law, legal validity, the nature of law and legal systems, punishment and responsibility, adjudication, legal reasoning, obligation and disobedience, professional ethics, epistemological issues involved in evidence and procedure, philosophical investigations of conceptual and foundational questions in law and legal systems, the impact of legal developments on philosophy, the relations of law and legal institutions to other social institutions, the relation of legal and political theory, the economic dimension of issues in jurisprudence and legal philosophy, the nature and development of legal concepts, and exploration of problems in particular areas of the law.

Subscription Information **ISSN** 0167–5249
1987, Vol. 6 (3 issues)
Institutional rate: Dfl. 212.00/US $89.00 incl. postage/handling
Private rate: Dfl. 82.00/US $34.00 incl. postage/handling
Private subscriptions should be sent direct to the publishers

Back Volume(s) Available
Volumes 1–5 (1982–1986

Price per Volume
excl. postage
Dfl. 175.00/US$62.00

D. Reidel Publishing Company

Kluwer academic publishers group

P.O. Box 989
3300 AZ Dordrecht
The Netherlands

101 Philip Drive
Norwell,
MA 02061, U.S.A.

Falcon House
Queen Square
Lancaster, LA1 1RN, U.K.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, Jane Heal, R. F. Holland, C. Hookway, C. A. Kirwan, D. H. Mellor, H. W. Noonan, M. A. Stewart, A. R. White, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices for 1988 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£13.00	£16.00	U.S. $33.00
Individuals	£8.00	£10.00	U.S. $21.00

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

CONTRIBUTIONS. Articles submitted for publication should be addressed to the Editor, Mr Peter Smith, Department of Philosophy, University College of Wales, Aberystwyth, Dyfed SY23 3DY.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

Vol. 48 No. 2 (New Series No. 218) March 1988

ANALYSIS

Edited by
PETER SMITH

CONTENTS

Rearrangement of particles: reply to Lowe	DAVID LEWIS
The problems of intrinsic change: rejoinder to Lewis	E. J. LOWE
Epistemic universalizability	BOB HALE
Faith, Hope and Charity: Russellian thoughts defended	GREGORY MCCULLOCH
More faith than hope: Russellian thoughts attacked	PETER CARRUTHERS
Carruthers repulsed	GREGORY MCCULLOCH
Privacy and verification	CHARLES B. DANIELS
Is modern moral scepticism essentially local?	RICHARD BETT
Some by the way remarks on Wreen's 'by' ways	KARL PFEIFER
A reply to Cahn	DANIEL H. COHEN
Ramsey–Lewis is better than Mackie	DAVID PAPINEAU

BASIL BLACKWELL · OXFORD AND NEW YORK

ISSN 0003-2638

ANALYSIS 48.2 MARCH 1988

REARRANGEMENT OF PARTICLES: REPLY TO LOWE

By DAVID LEWIS

I BACKGROUND

ORDINARY things, for instance we ourselves, undeniably persist through time. As we persist, we change. And not just in extrinsic ways, as when a child was born elsewhere and I became an uncle. We also change in our own intrinsic character, in the way we ourselves are, apart from our relationships to anything else. When I sit I'm bent, when I stand I'm straight. When I change my shape, that isn't a matter of my changing relationship to other things, or my relationship to other changing things. *I* do the changing, all by myself. Or so it seems. What happens must be possible. But how? Nothing can have the two incompatible shapes, bent and straight. How does having them at different times help? In *On the Plurality of Worlds* (Blackwell, 1986; henceforth *PoW*), p. 204, I listed three solutions, and said that only the third was tenable.

The first solution is that the 'properties' are really relations to times. That lets us say that things persist by *enduring*: the one thing is present at different times; and not mere temporal parts of it, different parts at different times, but *all* of it, wholly present at each of the times. The whole of me stands in the bent-at relation to some times and the straight-at relation to others. I complained that shapes are properties, not relations. No doubt a friend of the first solution will draw a distinction that he will *call* the distinction between matters of one's own intrinsic character and matters of one's relationships: having a shape will go on one side, being an uncle on the other. But call it what he will, his account reveals that really he treats shape, no less than unclehood, as a matter of relations. In this account, nothing just has a shape *simpliciter*. The temporary 'intrinsic properties' of things, so understood, do not deserve the name. This solution amounts to a denial that things really do have temporary intrinsics, and therefore is untenable.[1]

---

[1] The first solution has an 'adverbial' variant, defended in Sally Haslanger, *Change, Persistence, and Possibility* (Ph.D. dissertation, University of California, Berkeley, 1985); and, at least as a possibility, in Mark Johnston, 'Is There a Problem About Persistence?', *Aristotelian Society Supplementary Volume* 61 (1987), pp.

The second solution says that there is only one genuine time, the present. Intrinsic properties are genuine properties, and a thing can have them *simpliciter*, without regard to any relationships to anything else. However, the only intrinsic properties it has *simpliciter* are the properties it has *now*. What passes for persistence and change, on this solution, does not really involve other times. Rather, there are 'abstract' ersatz times, to go with the one 'concrete' genuine time. These represent, or misrepresent, the present. If I am bent now, and straight later, there is an abstract misrepresentation of the present according to which I am straight. 'Persistence' and 'change', so understood, do not deserve their names. This solution amounts to a denial of persistence and change, and therefore is untenable.[2]

The third solution, the tenable one, is that incompatible temporary intrinsic properties do not all belong to the same thing. A persisting thing *perdures*. It consists of temporal parts, or *stages*, different ones at different times, which differ in their intrinsic properties. When I sit and then stand, bent stages are followed by straight stages. Each stage has its shape *simpliciter*. Shape is truly intrinsic.

To be sure, my shapes belong in the first instance to my stages, and in a derivative, relational way to the whole of me. Persisting thing x is bent at time t iff some stage of x is at t and is bent. What distinguishes the first solution from the third is not that the third does away with shape-at-a-time relations. Rather, it is that the first has wrongly done away with shapes as intrinsic properties that can be had *simpliciter*.

Imagine trying to draw a picture of two different times, t_1 when I sit and t_2 when I stand. You draw two circles, overlapping because I exist at both times so you want to draw me in the inter-

---

107–35. The adverbial variant avoids my complaint that shapes are not relations. It puts the relationality not in the shapes themselves but in the having of them: there is a three-place relation of instantiation, this relation holds between me and bentness and some times, and it holds between me and straightness and other times. I ask: what does standing in some relation to straightness have to do with just plain being straight? And the variant still claims that to be shaped is to stand in relations to other things, *inter alia* to times. I say it still amounts to a denial that things have temporary intrinsics.

[2] The second solution also has an 'adverbial' variant, defended in discussion by Mark Hinchliff. It omits or plays down the ersatz times. Instead, temporal modifiers introduce relations of things to properties. If I am now bent, bent is just the way I am; whereas if I will be straight later, I stand in some relation to straightness which does not involve any having *simpliciter* of straightness by anything. Change between two times, one of them the present, is the modified-having of a property incompatible with a property the thing simply has. Change between two non-present times is modified-having of two incompatible properties. But the modified-having itself goes on in the present. I ask again: what does standing in some relation to straightness have to do with just plain being straight? And the variant still denies that 'persistence' and 'change' involve any genuine times except the present. I say it still amounts to a denial of persistence and change.

section. But then you have to draw me bent and also straight, which you can't do; and if *per impossibile* you could, you still wouldn't have done anything to connect the bentness to t_1 and the straightness to t_2 instead of *vice versa*. What to do? The first solution says to draw the circles overlapping, draw me in the intersection as a mere dot or shapeless blob, draw a line labelled 'bent-at' from me to the t_1 circle and a line labelled 'straight-at' from me to the t_2 circle. A queer way to draw a shape! The second solution says to draw only the t_1 circle, and in it draw me bent; and then draw not another circle but an abstract misrepresenter saying that I'm straight. The third solution says to draw two *non*overlapping circles, in the t_1 circle draw me bent, in the t_2 circle draw me straight, and then draw a connexion to indicate that the bent me plus the straight me add up to a single person.

II LOWE'S FOURTH SOLUTION

In his 'Lewis on Perdurance Versus Endurance' (ANALYSIS 47.3, June 1987, pp. 152–54), E. J. Lowe agrees that we need a solution, and joins me in rejecting the first and second solutions. But he rejects the third solution as well. He finds it 'scarcely intelligible' to say that things like people or puddles, as opposed to events or processes, have temporal parts. I disagree; but won't repeat here what I have said elsewhere about the intelligibility of temporal parts.[3] Lowe does find perdurance intelligible enough to be denied, and deny it he does.[4] After rejecting all three solutions, Lowe is urgently in need of a fourth.

I too would welcome a fourth solution, but for quite a different reason. If the third solution alone is tenable, then our commonsense belief in persisting things commits us implicitly to perdurance — and this despite the fact that some of us firmly reject the notion of temporal parts (except of events or processes) and many more have never heard of it! It would be better not to impute such surprising commitments to common sense, but only the plain commitment that things do somehow persist, never mind exactly how they do it.[5] If we had two tenable solutions — some sort of endurance theory, as well as the perdurance theory — to suspend judgement between, then such restraint would be feasible as well as sensible. Then we could leave the question of endurance versus perdurance forever unsettled, or perhaps settle it by finding out which coheres better with total science. At at any rate we could *not* settle it 'philosophically', by drawing out unsuspected consequences of common-sense intuitions. But if we're stuck with the third solution alone, restraint is beyond our reach.

---

[3] In my *Philosophical Papers*, Volume I (Oxford University Press, 1983), pp. 76–7.

[4] It is odd that when Lowe denies perdurance, he does not embrace endurance. I should have thought it automatic that if a persisting thing is present at a time, and not just partly present, then it must be wholly present. Lowe doesn't say why he finds 'wholly present at a time' problematic. Maybe he thinks the phrase carries a presupposition that the thing has, or at least is of a kind to have, temporal parts. I intend no such presupposition. If you think it is there, let it hereby be suspended. Then we may carry on saying that a thing 'endures' if it persists without having temporal parts; *a fortiori* if it persists and is of such a kind that it cannot intelligibly have temporal parts.

[5] Here I follow the lead of Johnston, op. cit.

Here is Lowe's fourth solution. Science teaches that things consist of particles. A change of shape for the thing, for instance for me when I sit and then stand, is a rearrangement of its particles. (Likewise for other intrinsic changes, for instance in my temperature or neural activity.) When the particles are rearranged, they undergo a change in their relations to one another; but no change in their intrinsic properties. In fact, it seems likely that fundamental particles *never* change their intrinsic properties. An electron or a quark has a certain charge, rest mass, and so on; all constant, from the creation of the particle to its destruction, no matter how the particle may move around and change its relations to other particles. There is no problem of intrinsic change for particles, if they have no temporary intrinsics. Particles, at least, may safely be supposed to endure; and larger things consist of these enduring particles, undergoing rearrangement but no intrinsic change.[6]

III FIRST OBJECTION, AND A REPLY

It may be true that particles have no temporary intrinsics. But it is far from certain, and far from necessary. Lowe's solution is at risk from scientific surprises. He might read in tomorrow's paper that the charges of electrons and quarks are not constant after all. They undergo fluctuations too small to detect with yesterday's instruments. If so, must Lowe conclude, after all, that there is no way he can persist? So it seems; because his solution to the problem of his own intrinsic change requires him to consist of particles not subject to intrinsic change.

Likewise, Lowe's solution makes no sense of counterfactual suppositions about persisting people in a world where particles fluctuate in charge. Such a world is possible — at any rate, counterfactual suppositions about it are intelligible — even if it contravenes the actual laws of nature. It doesn't matter whether the otherworldly fluctuating particles are genuine electrons and quarks, so long as some things made of these particles are genuine people — which they well might be.

---

[6] I wrote concerning 'fundamental particles *or momentary slices thereof*' (italics added) that 'maybe these things have no accidental intrinsic properties' and hence 'we would not face a problem of accidental intrinsics' for them (*PoW*, p. 205). Lowe cites this as conceding the point he needs; which is peculiar, because accidental intrinsics are not the same thing as temporary intrinsics. For instance a particle might well have the shape of its entire spatiotemporal trajectory as an *accidental*, but certainly not *temporary*, intrinsic property — wherefore my italicized phrase.

No harm done. Though the two hypotheses are independent, it is at least as probable that particles lack temporary intrinsics as it is that they, or their slices, lack accidental intrinsics.

I offer Lowe this reply. In the case of ordinary large things, such as ourselves, we are entitled to firm intuitions about what should count as just the way the thing itself is, and not at all a matter of relationships. For instance, we should be firm in resisting the idea that our shapes are really relations to times. But it is otherwise for the 'properties' of particles; even for the most familiar ones, charge and rest mass, let alone the new-found 'spins' and 'flavours' and 'colours'. It is up for grabs whether these are as intrinsic as we offhand think. Lacking close acquaintance with the submicroscopic realm, we should be open-minded. If charge turns out to fluctuate, and if our only solution to the problem of persistence requires that particles never undergo intrinsic change, that might be reason enough to conclude that charge — unlike shape! — is really a relation to times. Or if we suppose counterfactually that charge fluctuates, we might insist on making it part of the supposition that charge is not, after all, an intrinsic property.

IV Second Objection, and a Reply

Even if the problem of temporary intrinsic *properties* goes away, for particles at least, there is still a problem of temporary intrinsic *relations*.[7] Exactly as some properties are just a matter of how the thing itself is, without regard to any relationship to any second thing, so some relations are just a matter of how two things are *vis-à-vis* one another, without regard to any relationship to any third thing. The relation is intrinsic to the pair of *relata*. The ever-changing distances of particles from one another seem to be temporary intrinsic relations. Rearrangement of particles poses the same old problem of temporary intrinsics, except that now it is a problem about intrinsic relations rather than intrinsic properties. How can the same two things stand in different, incompatible intrinsic relations? How can the same two particles have two different distances? How does it help that they have them at different times?

We have three solutions, as before. First solution: particles have their temporary distances by standing in different relationships to different times. They endure, so whatever relations they enter into must be relations of the entire enduring particle. But there is no such thing as being a certain distance apart *simpliciter*. Instead there are three-place distance-at-a-time relations. This solution is untenable because it amounts to a denial that distances are intrinsic relations. Second solution: there is only one genuine time, the present, and the only distance there is between two particles is their distance *now*. What passes for persistence and change is the abstract misrepresentation of this one time, according to which the two particles exist and have some different distance. This solution is untenable because it amounts to a denial of persistence and change. Third solution: the particles perdure, and it is in the first instance their temporal parts that are various distances apart. Lowe will reject this as 'scarcely intelligible'. Properties or relations, it's all the same problem. What has Lowe gained?

---

[7] It corresponds to the problem of accidental external relations noted in *PoW*, pp. 205–6, just after the 'concession' Lowe cited. 'External' relations, in one sense of the term, are those intrinsic relations that do not merely supervene on the intrinsic characters of their *relata*.

Try again drawing the circles for t_1 and t_2, overlapping now because a certain two particles exist at both times and so you want to draw them in the intersection. They've undergone rearrangement, as particles often do. You have to draw them at two different distances apart, which you can't do; and if *per impossibile* you could, you still wouldn't have done anything to connect one distance to t_1 and the other to t_2 instead of *vice versa*. What to do?

I offer Lowe this reply. Suppose he is prepared to accept a substantival theory of spacetime, and one that distinguishes occupants of spacetime from the regions of spacetime they occupy. It might be nicer to stay neutral, but if Lowe has the problem about persistence that he thinks he has, a solution would be worth the commitment. Then for each particle, there is a spacetime region that is its trajectory. Lowe might grant that a spacetime region has temporal parts: it is like an event or process, unlike a person or puddle or particle. Suppose (for convenience only) that the particles are point-sized; then the smallest temporal parts of their trajectories are spacetime points. An enduring particle is wholly present at each point on its trajectory. Distances are intrinsic relations, sure enough; but of points, not particles. When two points are a certain distance apart, that's just a matter of how the two points are *vis-à-vis* one another, regardless of any relationship to any third thing. Enduring particles have their distances in a derivative and relational way: particles *x* and *y* are one metre apart at *t* iff points *p* and *q* are one metre apart, *p* and *q* are at *t*, *p* is one of the points where *x* is, and *q* is one of the points where *y* is. So in the first instance it is the trajectories that undergo rearrangement. They change their distances from one another in the unproblematic fashion of perdurers. Their different temporal parts, at different times, are at different distances. The particles undergo rearrangement vicariously: their trajectories do it for them. All the particle has to do is to occupy each point of its trajectory. That requires no change in intrinsic relations. The particle need only bear the same intrinsic relation of occupancy to each of many points.

Draw the circles overlapping; put the particles themselves in the intersection; but put all the spacetime points which the particles occupy outside the intersection, some in the t_1 circle and some in the t_2 circle. Draw lines of occupancy from the particles to their points.

This might be acceptable, as a last resort, if we had eliminated the alternatives. It is, peculiar to say that the particle is only indirectly involved in its own relationships of distance; but at least these turn out to be proper intrinsic relationships of *something*, and what's more, something intimately associated with the particle. What makes this picture of endurance barely acceptable is that it has so much *per*durance mixed in.

V FINAL OBJECTION

Waive the first two objections. Give Lowe enduring particles, and their rearrangement. I want to ask: what has this to do with *me*? My particles endure, but what do I do?

A strange worry. Surely I am nothing over and above my particles: I am them, they are me. The 'are' of composition is just the plural of the 'is' of identity. So if it's settled what they do, there can't be any further question what I do. You might as well say: I know all about the life of Cicero, now what about Tully?

I can agree to this, or near enough; but Lowe can't. Strictly speaking, of course, neither of us can say that I am my particles. If I were, I would exist whenever they do; but I don't. My beginning was a rearrangement of preexisting particles; my end will be a rearrangement of particles that will exist afterward. Further, I eat and excrete; so I cannot be identical both to the particles I was composed of yesterday and to those I am composed of today. But for me, as a perdurance theorist, the correction is close at hand. I consist of temporal segments of particles. So what's true to say, by way of identifying myself with my particles, is that I am my particle-segments. But for Lowe, who rejects particle-segments, this correction will be 'scarcely intelligible'. I think he has no other way to make sense of saying that I just *am* my particles. Composition as identity is not for him.

He will need, instead, to understand composition as a one-many relation of things that are in no sense identical. (It will be a temporary intrinsic relation, and so will give him a new problem of temporary intrinsics.) The relation of me to my particles is an interesting and intimate one, no doubt. But if it is not identity, then to say that my particles endure is simply not to address the question whether *I* endure; and how, if I do endure, I manage to undergo intrinsic change. What my particles do is, strictly speaking, irrelevant. Lowe has not offered any fourth solution, satisfactory or otherwise, to the problem of my intrinsic change. Rather, he has changed the subject.[8]

Princeton University,
Princeton, NJ 08544, U.S.A.

© DAVID LEWIS 1988

---

[8] I thank Donald Baxter, Sally Haslanger, Mark Hinchliff, and especially Mark Johnston for helpful discussions.

THE PROBLEMS OF INTRINSIC CHANGE: REJOINDER TO LEWIS

By E. J. LOWE

IN 'Rearrangement of Particles: Reply to Lowe' (above, pp. 65–72), David Lewis presents a formidable response to my article 'Lewis on Perdurance Versus Endurance' (ANALYSIS 47.3, June 1987, pp. 152–4). He concludes by saying that I have 'not offered any fourth solution, satisfactory or otherwise, to the problem of...intrinsic change. Rather, [Lowe] has changed the subject.' However, since writing my earlier article it has become apparent to me that there is more than one subject, because there are in fact two distinct problems of intrinsic change — a *semantic* problem and a *metaphysical* problem. Further, I believe that Lewis's favoured solution (in terms of temporal parts) fails not least because it attempts to cope with both problems at once, for this helps to make it a bad solution to each. It is a bad solution to the semantic problem because (and Lewis himself expressly recognizes this difficulty) it imputes surprising ontological commitments to the very meaning of our common-sense talk of persistence through change. But equally it is a bad solution to the metaphysical problem, in part precisely because it is one substantially driven by extraneous semantic objectives. These extraneous objectives also blind Lewis to the superior merits of the sort of scientifically motivated solution to the metaphysical problem that I defended in my earlier article.

The remainder of this paper is divided into two parts, the first dealing with the semantic problem of intrinsic change and the second with the metaphysical problem.

I

The *semantic* problem of intrinsic change is the problem of specifying the logical form of sentences ascribing temporary intrinsic properties to persisting objects, in such a way that we do not run into contradiction in describing such an object as undergoing a change from possessing one such property to possessing another incompatible one.[1] The problem may be seen as one of parsing a sentence of the unregimented form '*a* is *F* at *t*' in such a way that it is clearly compatible with one of the form '*a* is *G* at *t*'', even though '*F*' and '*G*' should express mutually incompatible intrinsic properties (as do 'straight' and 'bent'). Three solutions immediately suggest themselves. (These do not correspond exactly to Lewis's original three candidate solutions: the first and third correspond to his first and third, but the second corresponds to a solution mentioned in Lewis's first footnote.) We may parse '*a* is *F* at *t*' either as (i) '*a* is *F*-at-*t*' (so that what is really being ascribed to *a* is, say, a shape-at-a-time — a kind of *relational* property), or as (ii) '*a* is-at-*t* *F*' (so that the ascription of a shape to *a* is *temporally qualified*, i.e., the property-exemplification relation between *a* and a shape is relativized to a time), or finally as (iii) '*a*-at-*t* is *F*' (so that a shape is ascribed to a *temporal part* of *a* — Lewis's favoured solution).

It seems to me that if we have to accept one or other of these three solutions to the semantic problem of intrinsic change, then we had better plump for solution (ii), as this is clearly the least revisionary with respect to our common-sense talk of persistence through change. Solution (i) is revisionary about predicates and solution (iii) is revisionary about subjects, but solution (ii) retains '*a*' and '*F*' as subject and predicate respectively and takes 'at *t*' at its face value as having *adverbial* (or predicate modifier) status. I think it is tendentious for Lewis to describe solution (ii) as merely a 'variant' of solution (i), as I shall explain more fully in a moment. Nor am I persuaded by his reasons for finding solution (ii) problematical. On the latter score, he asks 'what does standing in some [temporally relativized] relation to straightness have to do with just plain being straight?'[2] But what does *Lewis* understand by 'just plain being straight'? And why does he attach such importance to the notion — why *should* there be anything to do with 'just plain being straight'? I can only assume that this phrase is supposed to denote some atemporal mode of property exemplification (such as is enjoyed, perhaps, by abstract geometrical figures). But in that case, why can't we just happily deny that changeable, persisting physical objects can ever correctly be described as 'just plain being straight'? It transpires, of course, that Lewis himself is committed to denying this too, since he holds that a changeable, persisting thing is only ever describable as 'being straight' in the derivative sense of having one or more straight temporal parts. Now it is true that the temporal parts in question supposedly *do* qualify as 'just plain being straight'.[3] But, given that Lewis and I can *agree* that a changeable, persisting thing is never describable as 'just plain being straight', what is it that compels one to accept *his* account of this fact, couched as it is in terms of entities that allegedly *are* so describable? Is it that the notion of 'just plain being straight' is supposed to have some sort of epistemic priority for us — that we have to learn first what it is for the temporal parts of things to 'just plain be straight' before we can get to grips with the more complicated situation with changeable, persisting things? That can hardly be so, given mankind's general ignorance of the very existence of these supposed temporal parts until a very few years ago! Surely it is on the contrary much more plausible to see the notion of 'just plain being straight' (if we really need it at all) merely as an abstraction from our temporal experience of the variable geometry of changeable, persisting things. Here I note that in criticism of solution (i) (and by implication of solution (ii) also) Lewis says that it 'has wrongly done away with shapes as intrinsic properties that can be had *simpliciter*' — even though it appears that by his own account the only (physical) things that *can* 'have shapes *simpliciter*' are *temporal parts*, since he concedes that changeable, persisting things cannot. But then I ask: given the widespread ignorance of the very existence of temporal parts, whence comes the supposed intuition that shapes *are* 'properties that can be had *simpliciter*'?

---

[1] Lewis is no doubt right in implying that an analogous problem arises for what he calls intrinsic *relations*: but since the solution I defend below can be extended to deal with them, I shall for simplicity restrict my discussion to intrinsic *properties*.

[2] As we shall shortly see, the words that I have interpolated here are not ones that Lewis would have chosen, but this is because he significantly misinterprets solution (ii).

Lewis says of solution (ii) that 'it still amounts to a denial that things have temporary intrinsics'. But I don't see this. The shapes and so on that things are by this account said to have are still treated as genuine *non-relational* properties, and only the *having* of them is related to a time — but, then, isn't that precisely what allows them to be *temporary* properties? Describing solution (ii), Lewis says 'It puts the relationality not in the shapes themselves but in the having of them' — and this is correct — but he then goes on to give the following gloss on this: 'there is a *three-place relation of instantiation* [holding between objects, properties and times]' (my emphasis). This gloss is however unacceptable. To say that the *having* (instantiating, exemplifying) of a shape is related to a time is to say that the holding of a *two*-place relation is related to a time, not that a *three*-place relation is involved, one of whose relata is a time. Lewis's gloss exhibits a failure to take seriously the adverbial (or predicate modifier) status of 'at *t*' according to solution (ii). Hence he is wrong to assert that this solution 'still claims that to be shaped is to stand in relations to other things, *inter alia* to times'. For what it in fact claims is that a thing's *being shaped* itself stands in relations to times, not that a thing's being shaped is partly a matter of *that thing's* standing in relations to times. Thus an adherent of solution (ii) will want to say, in explanation of the consistency of saying that an object *a* is bent at time *t* and straight at time *t′*, that this amounts to saying that *a*'s having a bent shape *obtains at t* while *a*'s having a straight shape *obtains at t′*.

---

[3] I say 'supposedly' because I cannot really see by what right Lewis can avoid having to say that a temporal part has an intrinsic property *at the time at which it exists*. The fact that it does not exist at other times and so cannot undergo intrinsic change seems beside the point. However, I shall waive this objection for the sake of argument.

Although one should exercise caution in drawing analogies between space and time, the following spatial analogy may be deemed illuminating. The analogy concerns the problem of what we might call '*local* intrinsics'. Consider the sentence 'The Thames is broad in London', which we plainly want to see as compatible with one like 'The Thames is narrow in the Cotswolds', even though 'broad' and 'narrow' express incompatible intrinsic properties. Again three ways of parsing our first sentence suggest themselves: (i) 'The Thames is broad-in-London' (ascribing to the Thames the relational property of broadness-in-London), (ii) 'The Thames is-in-London broad' (modifying the ascription of broadness to the Thames by an adverb of place which specifies *where* along its length the Thames is broad), and (iii) 'The-Thames-in-London is broad' (ascribing broadness to a spatial segment of the Thames). And again I would urge the superiority of solution (ii) — at least as an analysis of what one would normally be taken to *mean* by the original sentence. Someone uttering that sentence would *not*, I submit, normally be taken to be ascribing the esoteric relational property of broadness-in-London to the Thames, nor however to be ascribing broadness only to a spatial segment of the Thames rather than to the Thames itself. For observe here that verifying the sentence 'The Thames is broad in London' requires one only to individuate a certain *river* (the Thames) and a certain *place* (London) — and then to do some measuring of that river in that place. One is *not* required to individuate and then measure a certain *spatial segment of a river* (the-Thames-in-London) — indeed, one can be quite innocent of the thought that such entities as river-segments even exist. Where our analogy breaks down, of course, is in the fact that while it is philosophically uncontroversial to claim that extended objects like rivers do indeed *have* spatial parts — even if the common man may not be in the habit of individuating them — it is by no means uncontroversial to claim that persisting objects do indeed have *temporal* parts. But then this makes it all the more significant that even in the spatial case solution (iii) can and should be resisted, since if it should be resisted where the claim about parts is uncontroversial then all the more should it be resisted where the claim is not uncontroversial — for not to resist it here is, as Lewis realizes, to embroil the semantics of common-sense talk in metaphysical controversy.

II

The *metaphysical* problem of intrinsic change is the problem of how there can *be* objects for the description of which the semantic problem arises — that is, how there can exist objects such that we need to be able to say, without fear of contradicting ourselves, that *one and the same* such object may undergo a change from possessing one intrinsic property to possessing another incompatible one. My answer is that the identity over time of certain objects consists in the preservation of certain relationships between their constituents at any given time: thus the identity over time of a ship consists (roughly speaking) in the maintenance of certain structural relationships between the objects that constitute its component planks and spars at any given time, and the identity over time of a tree consists (again, roughly speaking) in the preservation of certain functional biological relationships between the objects that constitute its component roots, trunk, branches, leaves and so on at any given time. The reason why this view of the diachronic identity of such objects provides an answer to the metaphysical problem of intrinsic change is that it explains how, consistently with a tree (say) being numerically *one and the same tree* at times *t* and *t′*, it may none the less differ in respect of *some* of its properties at the two times — differ because *its* diachronic identity is consistent with a degree of replacement and/or rearrangement *amongst its components*, sufficient to allow for growth and maturation and so forth.[4] Thus one and the same tree may at one time be small and slender and at another tall and stout. And, of course, the same sort of story may be told about one of the components of a tree, such as a leaf, in terms of *its* components — though, as I explained in my earlier article, there is no reason to regard this regress as vicious, since it may be terminated at the level of fundamental particles.[5]

Now Lewis also has an answer to the metaphysical problem of intrinsic change, and indeed it is effectively the same as his answer to the semantic problem — though, of course, he fails to represent the two problems as genuinely distinct.[6] For him the identity over time of a tree consists in the obtaining of certain relationships between its supposed *temporal parts or stages* — relationships which are consistent with these temporal parts differing from one another in respect of at least *some* of their properties, such as their size and shape. Thus, for Lewis, the reason why there can be such a thing as a tree which changes from being small and slender to being tall and stout is that such a thing actually *consists* of a series of temporal parts or stages, one of which may be small and slender and another tall and stout. However, as a solution to the metaphysical problem this involves the quite gratuitous invocation of a category of entities hitherto unknown to layman and scientist alike — 'temporal parts'. I say 'gratuitous' because we already have to hand the solution which I outlined a moment ago — a solution only invoking entities either already familiar to the common man (things like the component roots and leaves of a tree) or else, as we proceed steadily towards microstructure, discoverable by science (things like cells and molecules and atoms). It is true enough that this solution is not also a solution to the semantic problem of intrinsic change, and indeed presupppposes such a solution. At bottom it is this, I think, that Lewis finds unsatisfactory about my metaphysical solution as presented in my original article. But then, as I have tried to make clear, the solution to the metaphysical problem has no business to be masquerading as a solution to the semantic problem, and so it is by no means to the credit of Lewis's own solution that it addresses both problems at once. The correct solution to the semantic problem (admittedly not tackled in my original article) is, I submit, solution (ii) of part I above.

---

[4] Lewis is right in imputing to me a denial that an object such as a human being or a tree is to be *identified* with its components at any given time — 'Composition as identity is not for [Lowe]'. But as far as the *metaphysical* problem of intrinsic change is concerned, this denial by no means condemns my solution to irrelevancy, as Lewis alleges. For my solution shows how changes in an object's intrinsic properties may supervene upon replacements and/or rearrangements amongst its components.

[5] I develop this view more fully in my 'Substance, Identity and Time', *Aristotelian Society Supplementary Volume* 62 (1988), in which I also criticize the temporal parts approach in greater depth. I might remark here that my view of diachronic identity is not one which Lewis's picturesque circle diagrams are well suited to illustrate — which may help to explain his evident puzzlement at my position. Equally alien to my approach is the way in which Lewis sets up the perdurance/endurance distinction.

It cannot be too much emphasized that we need to prise apart the semantic problem of intrinsic change and the metaphysical problem. If we do not, we are likely to end up with both bad semantics and bad metaphysics — and this, I believe, is precisely Lewis's predicament.[7]

University of Durham,
Durham DH1 3HN

© E. J. Lowe 1988

---

[6] More accurately, perhaps, he thinks that there is only a single problem and perceives it as being primarily a metaphysical one — certainly, the opening paragraph of his reply to me reads as the articulation of such a position. That Lewis sees 'the' problem as basically metaphysical also helps to explain his dismissive attitude towards solution (ii) of part I above — for this solution to the *semantic* problem manifestly leaves the metaphysical problem quite untouched (though I, of course, see this as a virtue rather than a defect).

[7] I am grateful to David Over for comments on an earlier draft.

EPISTEMIC UNIVERSALIZABILITY

By BOB HALE

THE best (most disturbing) sceptical arguments threaten not just knowledge but reasonable belief. My primary purpose will be to explain why one such argument — the 'argument from error' — fails to touch reasonable belief, even if it deprives us of knowledge; but I shall also question the argument's claim, at least in the formulation I shall be considering, to be effective against knowledge.

I confine attention to Jonathan Dancy's presentation.[1] The argument deploys an epistemological version of the well-known principle of universalizability in ethics, the principle that 'a judgement that an action is morally good is universalizable in the sense that by making such a judgement one commits oneself to holding that any relevantly similar action is morally good' (p. 12). Dancy explains that one action is relevantly similar to another (judged morally good, say) if the former has 'the properties which constituted one's reasons for the judgement' that the latter is morally good. And he adds — crucially here — that these properties must be ones whose presence or absence can be registered by the judge — they mustn't be 'evidence-transcendent', or 'such that they may be absent even though we have the best possible evidence for their presence', since 'a difference between the two actions which that person is unable to recognize cannot justify a difference in judgement' (ibid.). Dancy's sceptic now reasons as follows. Suppose I claim today to know that it will rain this afternoon, basing this claim upon 'the normal grounds', such as weather forecast, gathering clouds, etc. The sceptic observes that I have made just this sort of claim before, on similar grounds, and been wrong — it has failed to rain. Since there is no relevant difference between the grounds on which I now claim knowledge, and the grounds on which I made similar, but false, claims in the past, I ought to retract. I can't properly claim to know now, whilst admitting that I didn't know then, because 'the only fact which would justify such a difference in claim is one which is not available to me; facts about the afternoon's weather are evidence-transcendent in the morning' (p. 13).

Set aside the question whether this argument accomplishes its sceptical purpose as far as knowledge goes, and ask: Could it, or some adaptation of it, show that we don't even have reasonable beliefs? It is natural to suppose not. The argument seems to turn upon the truth-entailing character of knowledge. Whilst it is not open to us to maintain that, in the admittedly relevantly similar past cases, we did know that e.g. it would rain in the afternoon, it seems so far perfectly possible to insist that we did, in those cases, reasonably believe that it would rain, even though in the event it failed to do so. Dancy seems to concede this; or at least, he agrees that 'we cannot argue straightforwardly that a false belief cannot be justified' (p. 14). But he thinks there is a way of redeploying the argument that conclusion. The situation we are in, in cases of the sort in question, is one in which our (supposed) justification for our belief is such that we have, in the past, held just such beliefs on just such grounds, and sometimes our beliefs have been true, but sometimes false. But now, Dancy suggests, the sceptic is poised for the kill. Since we shall certainly have to agree that we have indeed held (relevantly) similar beliefs on (relevantly) similar grounds, he will insist that, if our present belief is to be justified, we ought to be able to cite some reason for thinking that the present case is of the former sort rather than the latter; that is, is one of those cases where our (supposedly justified) belief is true, rather than false. And that we cannot do. He can, that is, object that 'your belief is unjustified because nothing you can point to suggests that this is a case where your belief is true rather than one of the (admittedly rarer but still) indistinguishable (to you) cases where it is false'. (p. 14)

---

[1] *Introduction to Contemporary Epistemology* (Blackwell, 1985). All page references are to this book. For a somewhat different attempt to derive sceptical conclusions via epistemic universalizability, which I cannot discuss here, see Jonathan Adler 'Scepticism and Universalizability' *J. Phil.* 78 (1981). For a reply, see Anthony Brueckner 'Epistemic Universalizability Principles' *Phil. Studies* 46 (1984). Brueckner criticizes Dancy's use of the principle, on grounds other than those advanced here, in 'Losing Track of the Sceptic' *Analysis* 45.2, March 1985.

An immediate response would be to point out that it is, so far, totally unclear why we should not simply agree that we can point to no feature of our present case which likens it to those past cases in which our similarly grounded beliefs have been true, rather than those in which they have been false, but claim that this does not matter, since our beliefs were justified in *all* these cases. If the sceptic is to oblige us, via universalizability, to agree that our present belief is unjustified, he needs to establish that it is not relevantly different from other cases in which we must agree that our beliefs were not justified, and he has done nothing to establish that. In short, it seems that, granted that there is some acceptable principle of epistemic universalizability, Dancy's sceptic has simply failed to ensure that the conditions for its application are satisfied. I am not, of course, denying that pointing to relevantly similar past cases of false belief could put us under pressure to admit that our present belief is not, after all, justified. You might observe that in some such case, my belief was not in fact justified because there was some further fact, which I had simply overlooked, that undermined my supposed justification; you may then reasonably contend that, failing assurance that my present justification is not similarly flawed, my present belief is unwarranted. Or again, without pointing to any specific such flaw, you may observe that the proportion of similarly grounded *false* beliefs has been uncomfortably high. But in neither of these sorts of case does the mere fact that I have held similarly grounded false beliefs by itself generate pressure to revise our opinion about the reasonableness of my present belief.

That response is, I submit, enough by itself to dispose of this attempt to extend the argument from error to reasonable belief. But it is worth exposing a further difficulty, before considering the initial application of the argument to knowledge. Suppose that, having laid claim to reasonably believe that *p* on what I take to be adequate but non-conclusive grounds, I unwisely accept the sceptic's demand that I point to some reason to think that my present case of putatively reasonable belief is one of those cases in which my belief is true, and not one where it is false. What am I to say to him? Clearly it is no good my simply repeating my original grounds for believing that *p*, since they do not, *ex hypothesi*, differ relevantly from my grounds in cases where my corresponding belief proved false. The sceptic requires me to produce some new reason. If I had failed fully to indicate the grounds for my present belief, I might, on reflection, recognize that I had not said enough to distinguish them from grounds which I would judge insufficient to warrant belief, and I might then be able to articulate a relevant difference. Cases of this sort apart, it remains quite unclear how I could proceed. What is clear, however, is that I could only satisfy the sceptic by producing something which, either by itself or in conjunction with my original grounds, amounts to a *conclusive* reason to believe that *p*. Faced with strengthened, but still less than conclusive grounds, the sceptic would insist that, for all I've said, my present case might still be one where my supposedly reasonable belief is false, so that I still need some further reason for thinking otherwise. At least, that is what Dancy's sceptic ought to say, if he consistently applies the weapon Dancy offers him; it could hardly be of any philosophical significance whether I, or anyone else, has *in fact* held false beliefs on grounds of some particular kind. If the sceptic's challenge were legitimate at all, it would be legitimized by the bare possibility of someone's believing falsely that *p* on grounds of the kind of question.

If that is right, then, even if the sceptic had made his case, he would have succeeded only at the cost of irrelevance. For, still supposing the argument from error to deprive us of knowledge, it would be wholly unsurprising that it could be extended to undermine reasonable belief as well, by construing that notion as, in effect, truth-entailing. But no disturbing conclusion follows from that about reasonable belief, more reasonably construed. It might be claimed that this second objection simply mistakes the sceptic's real argumentative purpose, which is to show that the notion of reasonable belief, on non-conclusive grounds, is unstable. However, if that *is* the sceptic's purpose, his argument simply begs the question — to insist that someone equipped with non-conclusive grounds ought, *per impossibile*, to be able to point to some feature in virtue of which he justifiably takes the case to be one in which his grounds ground a true belief, not a false one, is to assume, not to argue, that non-conclusive grounds are insufficient.

Does the argument at least succeed against knowledge? I think not. The argument uses, and will not work without, a very strong universalizability principle. A comparatively weak principle would say that it can be the case e.g. that I did not know yesterday morning that it would rain that afternoon, but do know this morning that it will rain this afternoon, only if there is some relevant difference between the two cases. This principle is, perhaps, scarcely open to dispute, but it is far too weak to support anything like the argument from error. It could be, so far as this weak principle goes, that the only relevant difference between our two cases is that it didn't rain yesterday afternoon, but will rain this afternoon. Indeed, as stated, the principle allows that what makes the difference between cases of knowledge and non-knowledge may be some fact which is evidence-transcendent in a very strong sense — some fact which is in principle inaccessible to us at all times. But it is worth emphasizing that even if it were insisted — as it might seem reasonable to insist, and as anyone disposed towards anti-realism would insist — that relevant differences should not be evidence-transcendent in this strong sense, the fact that it didn't rain yesterday afternoon but will rain this afternoon could still constitute a relevant difference, since that fact is not strongly evidence-transcendent. Dancy's sceptic needs a much stronger principle, which rules that it can be the case that I didn't know yesterday morning that it would rain that afternoon, but do know this morning that it will rain this afternoon, only if there is some relevant difference in the evidence available to me at these times. This much stronger principle receives no support from general anti-realist considerations. It is difficult to see any other plausible motivation for it. Certainly it is, in general, totally implausible to insist that genuine differences between things must be detectable at all times at which they obtain. And it is hard to see any non-question-begging reason for thinking that knowledge is a special case.

What my inability to locate any relevant difference between the evidence I have this morning and the evidence I had yesterday morning *does* mean, of course, is that I can be justified in thinking, this morning, that it will rain this afternoon only if I was justified yesterday morning in thinking that it would rain that afternoon. And this in turn means that I cannot be justified in claiming knowledge now unless I would have been justified then in making a similar knowledge claim. That is, there is a somewhat different epistemic universalizability principle which we should (or so it seems to me) accept, which says, roughly, that a claim to know that *p*, made in circumstances C, can be justified, but a claim to know that *q* (some not relevantly different proposition), made in circumstances C′, not, only if there is some relevant and detectable difference between C and C′, e.g. in the evidence available in the two cases. This principle imposes a condition upon *claims* to knowledge which is near enough the same as that imposed upon knowledge itself by the unduly strong principle lately rejected. Dancy's own formulation of the argument, in terms of what we can or can't *claim* to know, suggests at least partial appreciation of the point that his strong principle at best imposes a constraint upon what we can justifiably claim to know, but does not point unequivocally to its acceptance — 'can claim' is here crucially ambiguous between 'can justifiably claim" and 'can truly claim'. If the foregoing argument is correct, he ought to mean the former, but I fear he means the latter.

When the strong principle is understood, as it should be, as regulating what we can justifiably claim to know, it will not sustain the argument from error. In our example, we need to distinguish three questions:

(1) Was I justified in claiming, yesterday morning, to know that it would rain that afternoon?
(2) Am I justified in claiming, this morning, that I knew yesterday morning that it would rain that afternoon?
(3) Am I justified in claiming, this morning, to know that it will rain this afternoon?

The strong principle constrains us not to return different answers to any pair of these questions unless we can point to some relevant difference in the circumstances in which the knowledge claims are made. The answers we should like to give are, of course, 'Yes' to the first and third and 'No' to the second. But we can perfectly well return those answers without flouting the strong principle. In the case envisaged, I have, this morning, decisive evidence that it did not rain yesterday afternoon. This requires a negative answer to (2). But it does not require a negative answer to (1), since whether or not I was justified in claiming, yesterday morning, to know that it would rain that afternoon does not depend upon evidence that only became available later — that question, unlike the question whether my claim was true, concerns the evidence available to me at the time. Nor does it require a negative answer to (3). For there is an obvious, and obviously relevant, difference: I have this morning decisive evidence that it did not rain yesterday afternoon — there is no reason why, in answering (2), I should restrict attention to evidence available to me yesterday morning — but I have now no comparable positive evidence that it will not rain this afternoon. I must, of course, admit that I might be wrong about this afternoon's weather, i.e. that it is consistent with my having just the grounds I do have for believing that it will rain that it should not rain — and I may have to admit that I have believed falsely that it will rain on just such grounds in the past. But this is not to the point. It would be so only if it had been already established that less than conclusive evidence is never sufficient to justify belief. That is what Dancy's extension of the argument from error tries to show, but, as I have argued above, that attempt is flawed. In the absence of independent argument for that claim, what enjoins a negative answer to (2) is not the fact that my yesterday morning's grounds were less than conclusive, but the fact that I now have ample evidence that I was wrong.

To summarize: there are two plausible principles of epistemic universalizability, which it is important to distinguish because one concerns the truth of claims to knowledge while the other concerns their justifiability, and claims to knowledge may be justifiable without being true. The former constrains us not to assign different truth-values to knowledge claims in the absence of any relevant difference. This may, plausibly, be understood as requiring relevant differences that are not absolutely evidence-transcendent, but there is no reason to require that they should reside in the evidence upon which the claims in question are based — or at least, Dancy's sceptic has provided no such reason. The latter constrains us not to hold that one knowledge claim is justified, another not, unless there is some relevant difference in the evidence upon which the two claims are based. Neither principle will sustain the argument from error, which relies upon a confused hybrid. The sceptic slides from thinking that what is a necessary condition for a correct differential judgement about the justifiability of different knowledge claims is a necessary condition for a correct differential judgement about their truth. The slide is illicit because, while a claim to knowledge can hardly be true without being justifiable, such a claim can be false without being unjustifiable. One thing that greases the slide is, I suspect, the ambiguity of e.g. 'can claim to know' between 'can truly claim ...and can justifiably claim...'. Another is a feature of universalizability in ethical contexts which might easily be supposed to carry over to epistemic ones, but does not do so. Assuming an agent not to have been culpably ignorant of any material fact, the judgement that he acted rightly, or wrongly, in a certain situation is based upon his state of information in that situation — the judgement is not reversed, if it subsequently transpires that there is some further fact of which he was excusably ignorant, but which is such that, had he known of it, he ought to have acted differently; we may, with hindsight, think that it would have been better, had he done so — but that is clearly another matter. Although believing is not, straightforwardly anyway, something we choose to do, an analogous point obviously holds for justified belief. But it does not hold for knowledge; as I have tried to argue, there need be no relevant difference between my non-knowledge yesterday morning and my knowledge this morning other than the fact that in the fact that I believed falsely yesterday morning but believe truly this morning.[2]

University of Lancaster,
Lancaster LA1 4YT

© BOB HALE 1988

---

[2] I am indebted to my colleague Jeremy Roxbee-Cox for helpful reactions to an earlier draft, and to the editor for some acute comments which have resulted, I hope, in a considerable improvement in the formulation of the argument of the last few pages.

FAITH, HOPE AND CHARITY: RUSSELLIAN THOUGHTS DEFENDED

By GREGORY MCCULLOCH

I

FAITH, Hope and Charity are as similar as you please but for some important details. They are physically very similar, are embedded in very similar environments, have very similar histories, and each seems to herself to be confronted by a dangerous cat. Each seems to think to herself a thought she would express thus: 'I'm going to kick that cat'. Each consequently lashes out. But here are the important differences. Faith and Hope confront distinct cats (Mildred and Consuela respectively), and Charity confronts no cat, but is hallucinating. Now some philosophers espouse the doctrine of Russellian thoughts (RT-doctine), the doctrine that a thought standardly expressible by e.g. 'I'm going to kick that cat' essentially concerns the cat (if there is one) to which 'that cat' would refer. And they mean this to imply (a) that Faith and Hope think different thoughts (since they think about different cats), and (b) that Charity thinks nothing of the corresponding type (since she confronts no cat). As I propose to understand it, the RT-doctrine is part of a claim that Intentional psychology as we normally conceive it is *externalist*, in the sense that seeings, believings and many other states of mind are taken in a relational or environment-involving way, as seeings of or believings about this or that environmental item. (For the general RT-doctrine see J. H. McDowell, 'On the sense and reference of a proper name', *Mind* 86, 1977, pp. 159–85. G. Evans applies it to demonstratives in 'Understanding Demonstratives' in H. Parret and J. Bouveresse, eds., *Meaning and Understanding*, de Gruyter, 1981, pp. 280–303.)

Some reject the RT-doctrine. With respect to the above case, their essential reason would be that our subjects are so similar: subjectively, things seem the same to each of them; objectively, they each lash out in the same manner because of how things seem. So (the thought continues) the lashings ought to succumb to parallel psychological explanations. Yet the Russellian thought theorists (RT-theorists) seem to rule this out. According to them, there are psychological asymmetries among our subjects, and so if the actions are explained in the usual way by the ascription, among other things, of thoughts entertained, the explanations will not be parallel. This is thought unsatisfactory for a variety of (not entirely independent) reasons. One may be a Cartesian and/or methodological solipsist, and think that matters psychological, when construed in a strict sense, cannot essentially concern extracranial items like cats. One may accept the RT-doctrine as part of ordinary psychology, but take the *Doppelgänger* examples to show that ordinary psychology is not the best psychology to be had. But some critics — a species of what I call 'restricted internalists' — intend something different again. This is the idea that, even if it is considered as part of ordinary psychology, and even if Cartesian or solipsistic assumptions are not made, the RT-doctrine is still to be rejected. My concern is with this sort of restricted internalism. First I give the best argument I can think of in its favour, then I suggest that, by itself, it is of little or no interest. (For the view see e.g. S. Blackburn, 'Thought and Things', *Aristotelian Society Supplementary Volume* 53, 1978, pp. 23–41, e.g. at p. 33; H. W. Noonan, 'Russellian Thoughts and Methodological Solipsism' in J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge University Press, 1986, pp. 67–90; and P. Carruthers, 'Russellian Thoughts', *Mind* 96, 1987, pp. 18–35.)

Restricted internalists do not deny that, in some sense of 'psychological', there are psychological asymmetries like those the RT-theorists posits. And after all, Faith kicked Mildred, while Hope kicked Consuela; and these are different types of behaviour, one of which may be illegal, the other not. Further, Charity kicked nothing; she just lashed out with her foot. So at this level of description of behaviour we have asymmetries, and (without Cartesianism and/or solipsism) there is no quick route to a complaint about the RT-doctrine's implication that asymmetrical explanations should be given. And note that even Charity's action can be explained, despite the absence of a Russellian thought: for she did hallucinate a dangerous cat. Neither does our restricted internalist object to the very idea of Russellian thoughts. After all, these can be represented by set-theoretic constructs containing the offending animals. Indeed, as 'restricted' implies, this type of internalism does not oppose externalism *per se*. Denying the RT-doctrine is one thing: insisting that all extracranial matters be eliminated from psychology is something else. So I shall focus almost exclusively on aspects of internalism which are merely opposed to the RT-doctrine.

Why should *anyone* oppose the RT-doctrine? Well, it cannot be excluded from *Doppelgänger* examples that, in some sense of 'behave', the subjects might behave in *precisely the same way*: they might start out in exactly similar postures, and their bodies might then move in just the same fashion. This can be made true by definition, since *Doppelgänger* considerations are thought-experiments. We imagine Faith who is ascribed a Russellian thought about Mildred. Then we imagine Hope or Charity, the same in all respects except for facts about the identity or presence of objects which affect the identity or presence of a Russellian thought. And there is no obvious bar to constructing the example so that these facts seem to have no essential bearing on how the subjects register the situation and react. Then it is offered as an evident truth that there should be some explanatory symmetry. But why should there be, given that we are not supposed to be making any Cartesian or solipsistic or revisionist assumptions? It will not do to take the normal course and appeal to orthodoxy, intuition, or standard practice in the theory of perception. The historical and theoretical background of these things is shot through with Cartesian and/or solipsism based on individualistic approaches to empirical psychology. Hence such appeals are likely to undermine what is distinctive about restricted internalism. Moreover, externalism too has sources of traditional support, notably in the later work of Wittgenstein, so it is only by ignoring (or taking a selective attitude towards) uncongenial features of the state of the art that one can hope to derive support from the aforementioned appeals. What is needed is proper argument against the specific kinds of externalism being rejected. In particular, we need argument against the RT-doctrine.

Rather than review the claims made by restricted internalists, I offer them a new argument. This is partly because their usual claims do not seem to me to avoid the criticisms just given. But the new argument also makes explicit an important thought underlying a general impulse towards internalism which ought to be exposed, and which, we shall see, seriously threatens strong versions of the RT-doctrine. All the more to that doctrine's credit, then, that it can, in a weaker version, resist this attack.

II

Here is the argument. (1) We can do many things, but we could do nothing without moving our bodies, and therefore there is a sort of bedrock of bodily movements to which Intentional explanations apply. I usher in a Golden Age, by irreversibly changing the political climate, by assassinating someone, by shooting her, by aiming a gun and firing, by straightening my arm and squeezing the trigger...like *this*. Here we have a number of things done: an ushering, a changing, an assassination, etc. And whether we discern many actions, or one action described in ways incorporating different aspects of its causal upshot, there remains the plausible claim that the series of descriptions will end with a description of the way a body or a part thereof was moved. Of course, we describe people as not moving a muscle, and go on (say) to blame them for *doing* this. But this can be set aside as a redundant case: people who do not move strictly speaking do nothing at all, although it is convenient to apply the vocabulary of action to them. Or we could modify our restricted internalism, and claim that for an undoubtedly central class of actions, explanatory symmetries are required, but ruled out by the RT-theorist. In any case, quibbling about inaction is not a promising line of defence for the RT-doctrine.

If descriptions of bodily movement are to figure as the observational basis for Intentional explanations, they must not be too precise. For instance, we must not envisage a geometrical taxonomy of the movements of limbs. I might say, of the above example: I bent my finger. But this does not describe the movement of my finger in the most accurate way possible, since many types of finger-movements are finger-bendings. Nor is it guaranteed that I could 'let my hand have its head' and reproduce the exact movement. More to the point, neither I nor my observers would know if I succeeded, since we are not that good at moving and observing our bodies. So fine-grained conceptions of movement have no place in Intentional psychology, and there is no behavioural bedrock in this sense. However, consider something weaker: for any pair of *Doppelgänger*-actions targeted at an object, there has to be (a) some relatively basic description of a relevant bodily movement, which (b) applies to both *Doppelgänger*. This claim is very plausible. Since Intentional explanations apply to actions, and actions (we can concede) involve bodily movements, it would be very odd if the purveyors of these explanations, who explain what is going on before their eyes, could not (re)describe these happenings in fairly basic bodily terms. This takes care of (a): and (b), as we saw, is got by definition. So the hope is to replace action descriptions like 'X ϕ-ed Y' with something like 'X ψ-ed', where ψ-ing is a relatively basic form of bodily movement which does not need a target object in order to be accomplished. In this way, claims (a) and (b) eliminate the offending objects from our explicanda.

(2) We now add the claim (c) that parallel Intentional-psychological explanations can be given of such pairs of *Doppelgänger*-actions. This flows from a reasonable demand on any sort of causal explanation. If, with respect to discipline E, phenomena x

and *y* fall under the same description *D*, then it is reasonable (*ceteris paribus*) to expect *E* to supply an explanatory description *D** which will apply to both cases. The expectation may be disappointed — similar effects can result from different types of cause — but it seems that the present sort of case is not of this frustrating type. Thus in the case of Faith, Hope and Charity, one might say that each of them *lashed out with her right foot*, and offer the explanation that they each *seemed to be confronted by a dangerous cat* (cf. Noonan, op. cit., pp. 68–70). And it is not clear why the restricted internalist should not be able to offer analogous explanations of *any* putatively object-targeted action: on the contrary, the invoked seemings etc. are themselves defined into the *Doppelgänger* examples. So why not make use of them?

Note an obscure phrase in the argument: it is said that the *Doppelgänger*-actions will admit of (re)description in 'relatively basic' bodily-movement terms. What is this? The issue is largely of a piece with the question of how restricted one's internalism is to be. Perhaps our restricted internalist could countenance quite rich descriptions of actions and the states of mind which produce them. Certainly, as we conduct our normal Intentional business, we do not see each other as lone body-movers driven by minimally effective states of mind. We see each other embedded in the world, charged with the Intentional effects of the environment, responding to its denizens and their features (cf. Noonan, p. 68). We see each other as assassins (not finger movers), seekers after taxis (not arm wavers), drinking companions (not right arm exercisers), philosophers (not pencil-pushers or key-tappers), lovers etc. And as is often pointed out in discussions of phenomenalism, the only generally available descriptions of *seemings* involve vocabulary apt for describing the world of material things. Just how restricted a form of internalism is to be deemed viable? (conversely, exactly what sorts of externalism are permissible — and why?) I shall not pause here, since this is not my problem *qua* RT-theorist. But with regard to the ban on the RT-theorist's invocation of particular objects, the idea is that although thought and action cannot in *general* be characterized in Intentional-psychological terms without externalist concessions, it is an extra step — to be resisted — to the RT-doctrine. If I act on object *x*, this is because I move my body in a certain way: this movement could exist just as it is without the eventual impact on *x*; and as we saw, there ought to be a description of this movement which makes no mention of *x*. Moreover, the *Doppelgänger* point cannot even be made unless we help ourselves to the seemings which involve a certain *type* of thing. If I see or seem to see *x*, then it is plausible to hold that I am struck as by a certain type of thing (which I may not find familiar), and hence it appears to follow that I *seem to see* something of this type; but — here is the rub — *x*'s identity (or even existence) is not essential to how things seem.

III

We now have a restricted internalism based on the claim that the relevant *Doppelgänger* will be behaviourally equivalent in some relatively basic way and that, relative to this way of behaving, they will succumb to parallel Intentional-psychological explanations. And if this is granted, there might seem to be no interesting theoretical role for Russellian thoughts to play. If the restricted internalist just does not bother to define them, nothing essential about (say) Faith's motivation will be lost, and everything essential about Hope and Charity, and about their relevant psychological similarities to Faith, will be retained. This certainly conflicts with any claim which the RT-theorist might make to the effect that Russellian thoughts have an *essential* explanatory role to play; for the opposing claim is precisely that it is always at least logically possible to explain behaviour without them.

I think the RT-theorist might as well concede this. Although some might hold out for the strong claim that Russellian thoughts are explanatorily essential, this is not incumbent on the RT-theorist, and — given the arguments just sketched — the claim looks vulnerable (see also Noonan, pp. 74–5). But this is no cause for the restricted internalists to celebrate, since their point is only, as it were, a theoretical one. For the RT-theorist's claim need only be that in standard Intentional situations like those involving our three subjects, standard Intentional practice does not consist in looking for suitably stark construals of the movements concerned, and then explaining these with the weakest Intentional attributions required. Instead, the claim continues, standard practice consists in seeing subjects straight off as reacting to the world around them — Mildred etc. — and in understanding these reactions in correspondingly object-invoking, externalist terms (e.g. 'she saw that that cat was dangerous and so she kicked it'). And who can deny this? As we have already seen — and as the restricted internalist does and must accept (on pain of relinquishing the epithet 'restricted') — we certainly do operate with externalist descriptions and explanations of actions, some of which, again no one can deny, involve the invocation of objects to which the RT-theorists so dramatically draw attention, and reflect in their conception of propositional content. So to the extent that the RT-theorist's aim is merely to help describe (with that same type of propositional content) a very pervasive feature of our Intentional state of being (i.e. of our minds as normally conceived), restricted internalism can be acknowledged then ignored. In general, this sort of externalist can reply to *any* internalist proposal that it incorporates a distorted picture of the Intentional scene: we hardly ever adopt the sort of descriptive and explanatory stance to which it points; we are not really *like that.*

Perhaps the internalist claim will now be pressed in something of a revisionist spirit: to the extent that Intentional psychology as normally conceived is object-invoking, then so much the worse for that. Look at all these potential explanatory symmetries it ignores! But what sort of a suggestion is this? If the idea is that we should just change our Intentional practices in the ordinary run of things, it is ridiculous. What possible reason could we have to start going about seeing each other (and ourselves?) in the appropriately stark terms? Why should we become such monsters? Civilization as we know and love it would crumble! In any case, as is very often pointed out, 'calibrating' each other with externalist states is unavoidable even in exclusively explanatory situations: on the one hand, because we do not have the right sort of knowledge of each other to proceed in any other way, and on the other, because such 'calibration' can be a useful source of information about the world at large. This goes just as much for thoughts about particular objects as for any others.

So what is the force of the theoretical possibility to which our restricted internalist points? Presumably this. In developing a rigorous system of psychological generalizations, where the aim is to treat like as like in so far as this is possible, so that the widest possible class of happenings can be subsumed under the one generalization, we should as far as possible eschew externalism and see our subjects starkly. And particular objects, as we saw, are especially vulnerable to elimination. But the general idea is just a modification of the demand that methodological solipsism be adopted as a research strategy in cognitive science, and thus at this point we start to fall back on the larger assumptions, mentioned at the beginning of this paper, which our restricted internalist was supposed to avoid. In any case, whether demands for moves towards internalism can be made reasonable, and, if so, whether that would be a reason for rejecting the RT-doctrine, are simply further matters altogether. Even if the demand for *full* solipsism is granted, why cannot externalism and internalism coexist, as it were, the one an approach to the mind as we actually find it, the other an approach to cognition as we hope one day to understand it? These are big questions, and while our restricted internalism can play a part in making us confront them, it is very small beer indeed when the confrontation gets under way (see my 'Scientism, Mind, and Meaning' in P. Pettit and J. McDowell (eds) *Subject, Thought, and Context*, Oxford University Press, 1986, pp. 59–94). Whereas Russellian thoughts still remain, as components of a way of understanding one topic of the confrontation: and my defence of them rests here.[1]

University of Nottingham,
Nottingham NG7 2RD

© GREGORY MCCULLOCH 1988

---

[1] I thank Peter Carruthers, Harold Noonan and Peter Smith for helpful comments.

MORE FAITH THAN HOPE: RUSSELLIAN THOUGHTS ATTACKED

By PETER CARRUTHERS

I

IN 'Faith, Hope and Charity: Russellian Thoughts Defended' (above, pp. 84–90) Gregory McCulloch takes as his main premiss that our folk-psychological notion of singular thought is Russellian. (In this context a thought is *Russellian*, whose identity and existence is tied to the identity and existence of the objects thought about; and a thought is *singular*, which is both of a sort apt to effect reference to an individual, and which has a cognitive content differing from that of any purely descriptive or general thought. Prime candidates for thoughts which are both singular and Russellian are demonstrative, such as Faith's thought 'That cat is dangerous'.) McCulloch's aim is first to provide an argument which would motivate the introduction of a non-Russellian notion of singular thought without drawing on any Cartesian or methodological solipsist premisses; but then to show that this concession to the opponents of Russellian thoughts is of little significance, resulting in a position which inhabits a kind of no man's land between the extreme revisionism of cognitive psychology on the one hand, and our folk-psychological theories and practices on the other.

The closest that McCulloch ever comes to *arguing* for the Russellian status of our folk-psychological notion of singular thought is his repeated emphasis upon the fact that individual objects are standardly mentioned in our everyday descriptions of such thoughts — as when we say that Faith thought Mildred to be dangerous. But this no more shows the individual objects to be essential to the thoughts described than the fact that we commonly recognize and identify other human beings on the basis of their appearance shows those appearances to be essential aspects of the people in question. For it may be that we tend to exploit reference to individuals in describing one another's thoughts merely for brevity or convenience (because precision may not matter very much), or because we do not know enough about the person's situation to describe their[1] thought more precisely. Or, as I argued in 'Russellian Thoughts' (*Mind*, 1987, pp. 18–35), it may be that such descriptions subserve purposes — such as acquiring information about items in the world — which have nothing whatever to do with folk-psychological explanation. But in any case it certainly does not follow that Faith's thought is of a type whose identity is tied to the existence of any particular cat.

---

[1] Here and throughout I use the colloquial plural pronoun in place of the pernicious masculine singular.

In this reply I aim to do two things: I shall argue that at least one of our common sense notions of singular thought is non-Russellian; and I shall offer a much stronger argument than McCulloch's, which is similar in motivating the introduction of a non-Russellian notion without drawing on any contentious theoretical assumptions, but which does so in a manner wholly subversive of his position.

II

First, however, some remarks about the proper use of the term 'folk psychology'. It became clear to me in the course of our correspondence that McCulloch is prepared to employ it very widely, to cover the full range of our common sense use of psychological notions. (See also his 'Scientism, Mind and Meaning' in P. Pettit and J. McDowell, eds, *Subject, Thought, and Context*, Oxford: OUP, 1986, pp. 59–94.) Yet this is, I believe, something of a departure from standard philosophical practice, where folk psychology is generally understood to be the interlocking network of beliefs — said to constitute a proto-scientific theory — which we employ in explaining and predicting the behaviour of intelligent agents. For in fact a considerable portion of our common sense use of terms like 'belief' and 'thought' forms no part of such an explanatory network. (On this, see my *Introducing Persons*, London: Croom Helm, New York: SUNY, 1986, and my 'Conceptual Pragmatism', forthcoming in *Synthese*.) Nor is there anything to be gained from McCulloch's extended usage. On the contrary, it fosters the illusion that common sense constitutes a unitary perspective, and that that our ordinary notion of thought is unambiguous.

As I argued in 'Russellian Thoughts', there is certainly a standpoint within common sense which can motivate a Russellian notion of thought. However, this would not be the sort of hybrid thought espoused by contemporary Russellians like McCulloch, which consists of a Fregean sense (a mode of presentation) essentially tied to the existence of the object thought about, but rather the pure form adopted by Russell himself, which would consist, in the simplest case, of an ordered pair of an individual and a property. The standpoint in question is *belief acquisition*. Often what interests us, in receiving descriptions of someone's thoughts, is that we might in consequence have reason to make alterations within our own stock of beliefs about the world. In such cases, all that matters is which individuals the subject's thoughts concern, and which properties their thoughts ascribe to them. Since their modes of thinking about those objects and properties are of no importance, it is thoughts described from this perspective which are *de re*. But the notion of singular thought which results is too coarse-grained to be of any explanatory significance. For example, someone faced by a number of cats in a room full of mirrors may think 'That one is dangerous and that one is not' with respect to what is, in fact, the very same animal, and may take what seems to themselves to be appropriate avoiding action. Since on the pure Russellian view, the thinker in this case entertains both a singular thought and its negation, their action will not be rationally explicable (or at least not by attributing such a thought to them).

Of course besides the belief-acquisitor's perspective, common sense also includes folk psychology proper. And as I shall argue shortly, when we adopt this explanatory standpoint the appropriate notion of singular thought (and the notion which we do in fact employ) is non-Russellian. In my view the sort of hybrid Russellianism adopted by McCulloch and others results from a failure to distinguish between these two perspectives on thought description, and from choosing examples which cross different features of them. (Common sense also contains perspectives not relevant to the present debate, such as the one I refer to in 'Conceptual Pragmatism' as 'The Standpoint of the Practical Reasoner'. This is adopted, for example, by anyone wishing to persuade a thinker to take an alternative course of action to the one they propose.)

III

My claim in this section is that folk psychology (narrowly conceived of) is in fact non-Russellian. Consider first an example where the context is clearly explanatory — where there is no question of our interest in descriptions of thoughts being to acquire new information about the world — and where there is no lack of knowledge of the precise nature of the subject's mode of thinking. Suppose that Faith and Hope are engaged by a toy manufacturer to provide a measure of the degree of life-likeness in the appearance of their products. Each sits in a similar cubicle and is presented with exactly similar (but numerically distinct) toys at the same time, her task being to record her impressions on a 'realism scale' of 0 to 10. Generally the toys are pretty unconvincing. But suppose one day that when a new type of clockwork mouse is placed in front of her, scuttling across her desk as if genuinely alive, each thinks 'That mouse really does look real', and momentarily jumps up out of her chair in alarm. And suppose further that they both tell us about this afterwards. If we were asked to explain why they had behaved so similarly, should we not naturally be inclined to say 'Because they both happened to think the very same thought at that moment'? We should not be in the least tempted to say that they had both entertained thoughts belonging to different, though similar, types. Yet if the notion of thought in play were Russellian, this is just what we should have to say, since the toys in question were distinct individuals.

Consider next the case of Charity who, hallucinating the presence of a cat, essays the thought 'That cat is dangerous'. From the perspective represented by folk psychology it is hugely counter-intuitive to suggest, as the Russellian does, that Charity here entertains no singular thought — even conceding the presence of various descriptive beliefs, such as that there is *a* dangerous cat in front of her, which may be sufficient to explain her behaviour. Rather, our feeling is that Charity's behaviour is to be explained in precisely the way that the similar actions of Faith and Hope are to be explained, by the motivating presence of a singular thought.

Of course these examples by no means constitute a conclusive case. For one thing, it is quite easy to produce examples where our intuitions pull in a Russellian direction. But I maintain that they are all ones in which our interest is not (or not exclusively) explanatory, but where our standpoint is rather that of an acquisitor of new beliefs about the world. For another thing, there is the argument which has been used by many Russellians, that in a case like that of Charity there is no way to specify the content of her supposed singular thought. (See for example Gareth Evans *The Varieties of Reference*, Oxford: OUP, 1982, pp. 82, 139–40, 170 and 173.) But this is not an argument which can be available to McCulloch, since he allows the intelligibility of a non-Russellian motion of singular thought; and I have in any case rebutted that argument — to my own satisfaction at least — in 'Russellian Thoughts'.

IV

Having argued (albeit sketchily) that folk psychology is in fact non-Russellian, I shall now claim in this section that it is *right* to be so. The question is whether non-Russellianism can be shown to be well motivated, given that it commits us to providing dual-component explanations of action. Thus our explanation of the fact that Faith kicked Mildred whereas Hope kicked Consuela would have, strictly speaking, to split into two distinct parts: the statement that each entertained the very same thought, namely a demonstrative thought to the effect that the cat which they took to be in front of them was dangerous; and the statement that Faith was stationed in front of Mildred whereas Hope was placed in front of Consuela. Would it not be somewhat smoother to explain their actions as the Russellian does, saying simply that Faith thought that Mildred was dangerous, whereas Hope thought that Consuela was dangerous?

This brings me to McCulloch's argument that it can indeed be reasonable to introduce a non-Russellian notion of thought, given the project of explaining actions described in suitably minimalist vocabulary — foot-lashings, arm-wavings, finger-bendings and the like. Since he is clearly correct that such descriptions take us a long way from folk psychology, it is incumbent on me to find an argument for the same conclusion which employs our ordinary full-bloodied descriptions of action. And in fact such an argument is not far to seek. For McCulloch has merely chosen an example congenial to his purposes. It is because the only act in common between Faith, Hope and Charity is of the most minimal sort (a foot-lashing), whose description would hardly ever figure in our normal explanatory practice, that he is able to argue that it is only reasonable to explain the behaviour of all three in terms of the possession of a common thought if our project is to explain actions described at this basic level.

Consider an example which begins as McCulloch's does, with Faith and Hope each confronted by a different dangerous-looking cat, and with Charity confronted by no cat at all since she is hallucinating. But suppose that instead of lashing out at the cat with their feet, the story proceeds as follows (conforming, of course, to all the usual *Doppelgänger* restrictions): each thinks to herself 'That cat is dangerous', and consequently sets out to phone the police and the RSPCA, to warn her neighbours, and (eventually) to create a group of cat-capturing vigilantes, dedicated to keeping cats like that off the streets. At each stage as the story unfolds their actions are guided by memory-based singular thoughts — 'I must warn my neighbours about that cat', 'If the authorities won't protect us from cats like that one, then we shall have to defend ourselves' and so on. So the ultimate explanation of the whole story is to be found in the original episode.

Here then is an example where the Russellian is committed to giving different explanations of actions which are similarly characterized, even from an ordinary folk-psychological point of view — 'Telephoning the police', 'Warning the neighbours of a dangerous cat', 'Purchasing a large net', 'Forming a vigilante group' and so on. For in the cases of Faith and of Hope the explanations would lie in their singular (Russellian) thoughts. But Charity's otherwise similar actions would have to be explained in terms of her descriptive (non-singular) beliefs. We are therefore provided with just the sort of argument for employing a non-Russellian notion of thought which McCulloch himself gives, but in this case premised upon our ordinary folk-psychological descriptions of action, as opposed to revisionary minimalism. For not only would such a divergence in explanation be intuitively highly implausible, but like McCulloch we can argue that, other things being equal, similar effects are best explained by similar causes — the similar cause, in this case, being the act of thinking the very same (non-Russellian) singular thought.

A normal Russellian would be able to shrug off the above argument. In so far as they would claim it to be impossible to characterize the content of the singular thought which Charity is supposed to entertain, they would say that we cannot in any case achieve parity of explanation in these examples. In which case the argument for the explanatory superiority of the non-Russellian notion collapses. But this is not an option which is available to McCulloch, since he is prepared to allow a non-Russellian notion of singular thought to be intelligible. Thus, so far from having made only a minimal concession to the opponents of Russellianism, as McCulloch intends, he has in fact cut the ground from under his own feet. The moral to be drawn is that to allege the unintelligibility of singular thought in a case of hallucination like that of Charity is by no means optional for a Russellian, and that McCulloch's attempted weak-Russellian compromise is unstable.

To conclude: McCulloch is wrong to claim that our ordinary explanatory practice employs a Russellian notion of singular thought, and wrong in supposing that the only motivation for introducing a non-Russellian notion (which would not depend upon Cartesian or solipsist assumptions), would have to involve a commitment to minimalist descriptions of action.[2]

University of Essex,
Colchester CO4 3SQ

© PETER CARRUTHERS 1988

---

[2] I am grateful to Gregory McCulloch, David Smith and Peter Smith for their comments on an earlier draft.

CARRUTHERS REPULSED

By GREGORY MCCULLOCH

AS interpreted by RT-theorists, the reports

(1) Faith thinks that that cat [viz. Mildred] is dangerous,

and

(2) Hope thinks that that cat [viz. Consuela] is dangerous,

ascribe different thoughts to their respective subjects, since Mildred and Consuela are distinct individuals. But Faith and Hope could be *Doppelgänger*, and some say that subsequent actions of Faith and Hope which would be explained by (1) and (2) respectively ought therefore to succumb to parallel psychological explanations: something that the RT-theorists' interpretation apparently rules out. In 'Faith, Hope and Charity: Russellian Thoughts Defended' (above, pp. 84–90), I set myself the very limited task of defending the RT-doctrine against a version of this argument which would be available to certain 'restricted internalists', that is those of its opponents who profess not to make any Cartesian and/or solipsistic assumptions about the nature of the mind. My strategy was first to note that, in the example I chose, where Faith and Hope each kick the appropriate offending cat, the call for explanatory symmetry cannot immediately be made out, since the actions to be explained are themselves of a different type: Faith *kicked Mildred*, Hope *kicked Consuela.* However, I was prepared to concede that, given certain plausible claims about the role of (relatively) basic bodily movements in action on objects, the demand for symmetry could be upheld, even in the absence of the aforementioned forbidden assumptions. For example, one could ask for parallel psychological explanations of why both Faith and Hope *lashed out with a right foot.* I then went on to suggest how this demand might be met; one might offer the explanation that both Faith and Hope *seemed to be confronted by a dangerous cat.* But these concessions, I argued, need not trouble certain RT-theorists, whose interests lie in reflecting, in their theory of thought, the externalist nature of folk-psychological explanation. For in my example both explanans and explanandum employ minimalist conceptions of action and mentality which are alien to folk-psychology's externalist nature.

This was not intended to be an argument for the conclusion that our folk-psychological conception of singular thought is Russellian. The point was merely to emphasize that the RT-doctrine accords well with normal folk-psychological practice, whereas the minimalist stance which, I suggested, might plausibly underpin the restricted internalists' use of the *Doppelgänger* argument, does not. It is, of course, standard practice for internalists generally to try to *explain away* the externalist nature of folk-psychological ascriptions in a manner intended to leave room for the claim that the thoughts ascribed in the likes of (1) and (2) are non-Russellian: but this was not my concern. Rather, I was concerned to undermine any *motivation* which the *Doppelgänger* argument might give for this manoeuvre, even in the absence of the forbidden assumptions. It seems to me to be very important to approach these matters without tacitly making the Cartesian assumption that the onus is very much on Russellians to make out their outlandish doctrine in the face of revealed wisdom about the nature of the mind.

The main burden of Peter Carruthers's reply, however, is that the concessions I made to restricted internalists are 'wholly subversive of [the relevant RT-theorists'] position', since they can be made to yield a demand for parallel psychological explanation for actions as described in an externalist, rather than a minimalist, fashion (see 'More Faith than Hope: Russellian Thoughts Attacked', above, pp. 91–6; the quoted phrase is from p. 92). But it does not seem to me that Carruthers has established his conclusion.

In the first place, Carruthers's argument is based upon a misapprehension about what I conceded. He repeatedly says that I allow the intelligibility of 'a non-Russellian notion of singular thought', according to which the thoughts ascribed in (1) and (2) could be identical even though Mildred and Consuela are distinct individuals. But I do no such thing. I merely concede that there is *some* psychological ascription (viz. *seemed to be confronted by a dangerous cat*), which is both common to Faith and Hope and capable of explaining their (suitably described) subsequent actions. And it needs further argument to get from this to the claim which Carruthers attributes to me: argument which I neither endorse in my original paper, nor would wish to. It is, for example, very implausible indeed to suppose that the full nature of thoughts about objects — even objects which are given in sense experience — can be explained wholly in terms of how things experientially seem to the thinkers of the thoughts. Thinking and experiencing are different kinds of thing, as critics of empiricism know. So in speaking of the seemings I make no general commitment about the nature of demonstrative thoughts.

Of course, in allowing the seemings, I allow that at least some explanations of action in terms of thoughts are, in a certain sense, redundant. But my aim never was to deny this, or to claim — as Carruthers seems to imply that I do — that explanation in terms of (Russellian) thoughts is preferable to that delivered by (say) dual component approaches, because it is 'somewhat smoother'. I merely point out that explanations of bodily movements in terms of seemings 'take us a long way from folk psychology', as Carruthers nicely puts it, so that allowing them does not, *without further anti-Russellian argument*, undermine the RT-theorist I set out to defend. But the only place in Carruthers's reply where I can find such further argument is in his section III, where he invites us to consider what we would say about the identity of thoughts in certain imagined situations involving my *Doppelgänger*. However, as I pointed out in my original paper, such appeals to intuition are as likely as not to call up judgements based on prior acceptance of the Cartesian and/or solipsistic assumptions which the restricted internalists are supposed to avoid, and so are of especially little value in the present context. In any case, I myself do not feel disposed to say what Carruthers claims 'we' should say about his little scenarios.

Overall, then, it seems to me neither here nor there that, in Carruthers's main example, the actions to be explained are characterized in an externalist manner (after the initial episode, each of Faith and Hope 'sets out to phone the police and the RSPCA, to warn her neighbours, and (eventually) to create a group of cat-capturing vigilantes dedicated to keeping cats like that off the streets'). The common 'ultimate explanation' which I would allow the opponents of the RT-doctrine, and which proceeds in terms of the posited seemings, itself remains a long way from folk psychology. Still, the hostile reader may feel dissatisfied: *what about* the fact that the actions to be explained are described in an externalistic manner? Carruthers's example will certainly bear some further scrutiny.

The demand for explanatory symmetry arises, we are told, because Faith and Hope respond to the original episode in the same way. That is, their subsequent careers can be subsumed under the same (externalist, folk-psychological) descriptions: and, as Carruthers points out, I allow in my paper that *ceteris paribus* it is reasonable to look for similar causes of similar effects. Now it is not clear that the example Carruthers gives escapes the obvious reply that Faith and Hope in fact perform different types of actions in their subsequent careers. Do they live in the same district, have the same neighbours, notify the same police force, purchase the same net, and so on? They had better, obviously, otherwise the same reply is available to the RT-theorist here as was available in the original episode I described, where Faith and Hope perform different types of action since they kick different cats. In any case, it is very important to realize that the fact that two careers *can* be subsumed under the same folk-psychological description does not by itself motivate a call for parallel explanations. For example, at some time both Carruthers and I *set out to become professional philosophers*. But it does not seem at all plausible to claim that there should be some parallel folk-psychological explanation for this (we are both very perverse individuals?). At best, there is perhaps a research project here for an ambitious sociologist. Furthermore, something similar applies even in cases where actions are targeted on a particular object. For example, in July both Carruthers and I *went to Cambridge to attend the Joint Session*; but again, the call for parallel psychological explanations sounds distinctly hollow. Thus I must admit to a certain looseness of formulation in my claim about similar effects justifying, *ceteris paribus*, the search for similar causes: the '*ceteris paribus*' clause is evidently doing a lot of work. However, Carruthers and other restricted internalists can hardly draw comfort from this admission, since the loose formulation was made in the course of a charitable consideration of their own position, which I reject anyway.

Then why should we automatically be disposed to think, even when eschewing the forbidden assumptions, that the subsequent careers of Faith and Hope *might well* motivate a demand for parallel psychological treatment? One thought, clearly, is the idea — which can be defined into the example — that things seem very much the same from the point of view of each *Doppelgänger* as their careers unfold. As for the rest, I suggested this:

> . . . it cannot be excluded from *Doppelgänger* examples that, in some sense of 'behave', the subjects might behave in *precisely the same way*: they might start out in exactly similar postures, and their bodies might then move in just the same fashion. (Above, p. 86)[1]

But this, of course, takes us back to my original argument — and its conclusion.

University of Nottingham,
Nottingham NG7 2RD

© GREGORY MCCULLOCH 1988

---

[1] Note too that this is precisely the conception of 'exactly similar action' with which Carruthers operates in his 'Russellian Thoughts', *Mind* 96 (1987), pp. 18–35: see, for example, p. 33.

PRIVACY AND VERIFICATION

By CHARLES B. DANIELS

A PROPOSITION is *publicly verifiable* if more than one person can verify it. I wish to show the following: if some true proposition is verified by exactly one person, there is a true proposition that is not publicly verifi*able*.[1] Symbolizing 'x sometime verifies that p' as 'Vxp', I shall establish that if some truth A is privately verified by a person M, then for some truth A′ it is not possible for anyone but M to verify that A′.

Take a coin,[2] put it in a tin box, cover it tightly, and shake the box. Exactly one of the following propositions is true at time t:

---

[1] The formal proof I present mimics one constructed long ago by Frederic Fitch. Fitch proved that if some true proposition is unverified, some true proposition is unverifiable. I first heard the proof from Fitch over lunch at Yale in 1966. It has surfaced recently as the subject of discussion in: Dorothy Edgington, 'The Paradox of Knowability', *Mind*, 94 (1985) 557–68; Timothy Williamson, 'On Knowledge and the Unknowable', *Analysis*, 47 (1987), 154–8; 'On the Paradox of Knowability', *Mind* 96 (1987), 255–61.

[2] Or a flat beetle.

B. The coin is heads-up at t.
C. It is not the case that the coin is heads-up at t.

Have a dying man, Mr M(utely) M(oribund), look into the box to verify whether B or C is true. His last act is to replace the cover on the box and shake it well. Provided that there is no omniscient God or the like, whether B or C is true will never in fact be verified by anybody but M. So one of the following is also true:

D. $B \wedge (\forall z)(VzB \supset z = M)$
E. $C \wedge (\forall z)(VzC \supset z = M)$

Now it can be shown that no proposition of the form '$q \wedge (\forall z)(Vzq \supset z = M)$' can be verified by anyone other than M. Thus either D or E is true and not publicly verifiable. The proof is straightforward:

1.	$\Box$ $(\exists x)(x \neq M \wedge Vx(q \wedge (\forall z)(Vzq \supset Z = M)))$	Hyp
2.	$a \neq M \wedge Va(q \wedge (\forall z)(Vzq \supset z = M))$	Hyp
3.	$a \neq M$	$\wedge$-out,2
4.	$Va(q \wedge (\forall z)(Vzq \supset z = M))$	$\wedge$-out,2
5.	Vaq	V1,4
6.	$q \wedge (\forall z)(Vzq \supset z = M)$	V2,4
7.	$(\forall z)(Vzq \supset z = M)$	$\wedge$-out,6
8.	$Vaq \supset a = M$	$\forall$-out,7
9.	$a = M$	$\supset$-out,5,8
10.	$\sim(p \wedge \sim p)$	Hyp
11.	$a = M \wedge a \neq M$	$\wedge$-in,3,9
12.	$p \wedge \sim p$	$\sim$-out,10,11
13.	$p \wedge \sim p$	$\exists$-out,1,2,12
14.	$\sim(\exists x)(x \neq m \wedge Vx(q \wedge (\forall z)(Vzq \supset z = M)))$	$\sim$-in,1,13
15.	$\Box \sim(\exists x)(x \neq M \wedge Vx(q \wedge (\forall z)(Vzq \supset z = M)))$	$\Box$-in,1,14
16.	$\sim\Diamond(\exists z)(x \neq M \wedge Vx(q \wedge (\forall z)(Vzq \supset z = M)))$	$\Diamond$-df,15

All the steps of this proof are legitimate in first-order modal logic save 5 and 6. The rule V1 is relatively uncontroversial for verification: if x verifies that $(p \wedge q)$, x verifies that p. The rule V2 is even less controversial: if x verifies that p, then p. We can, therefore, draw the following conclusions:

(1) If some truth is verified by at most one person, then some truth is verifiable[3] by at most one person.
(2) If all truths are verifiable by more than one person, then all truths are verified by more than one person.
(3) If all truths are publicly verifiable, then all truths are publicly verified.[4]

---

[3] Verifiability is *possibility of verifying*, and it should be strongly emphasized that the notion of possibility involved here is *very* weak, logical possibility.

[4] One wonders whether 'the private language argument', whatever the argument is, validates the antecedent of (3).

A similar proof can be constructed for truths that are verified by no more than n people. We are entitled to conclude, then, that:

(4) If some truth is verified by at most n persons, then some truth is verifiable by at most n persons.

This is to say:

(5) If all truths are verifiable by more than n persons, then all truths are verified by more than n persons.

Let n + 1 be the number of persons in the universe over the fullness of time. We have then:

(6) If all truths are universally verifiable (verifiable by everyone), then all truths are universally verified.

Finally, since verification suffices for knowledge, we can conclude that:

(7) *If all truths are universally verifiable, then everyone is omniscient.*

University of Victoria,
P.O. Box 1700, Victoria,
British Columbia, Canada V8W 2Y2

© CHARLES B. DANIELS 1988

IS MODERN MORAL SCEPTICISM ESSENTIALLY LOCAL?

By RICHARD BETT

MY title is taken from a paper by Julia Annas, in which the following claim is made. 'Modern moral scepticism...is essentially local, a part of a globally unsceptical world-view which is likely to be scientifically based.' ([1], p. 17) The suggestion is that, if one is to be sceptical about moral values, one cannot at the same time subscribe to any form of scepticism about the world in general. The view that is here labelled 'modern moral scepticism' is the view that moral values are not objective, are not part of the world as it in itself (but are merely inventions or projections of ours); but to believe this, it is claimed, one needs to believe that we can refer to *other* items which *do* constitute a world 'as it is in itself'. Obvious candidates for the role of items which belong to the world as it is in itself are the entities described by the hard sciences; and so the modern moral sceptic will typically be a staunch objectivist about protons, electrons and the like. But at any rate, there must be *some* conception of the world about which the moral sceptic is staunchly objectivist. Otherwise, the argument goes, there is nothing with which moral values are contrasted, and so nothing in comparison with which their status can be downgraded. Thus, if one ceases to be an objectivist about science, one no longer has any reason to be sceptical about moral values, because moral values will no longer present themselves as unreal by comparison with scientific entities. More generally, if one does not think that *any* of our ways of talking about the world can describe it 'as is in itself', one will have no grounds for moral scepticism; for if *everything* that we talk about is in some sense constituted by us — in that thought and talk themselves inevitably import some set of categories, some perspective, some 'paradigm', some conceptual scheme, or whatever one takes it to be — then moral values, constituted though they may be by us, will have no less right to be regarded as part of the world than anything else. One cannot be a moral sceptic, then, unless one is healthily unsceptical about topics other than morality.

Annas's particular purpose, in presenting this picture, is to contrast modern moral scepticism with ancient moral scepticism, which was just one aspect of a broader sceptical attitude which applied to all beliefs about all subjects. However, the similarities and differences between ancient and modern scepticism are not my concern; what interests me here is simply the specific view about modern moral scepticism which I have just sketched. The reason why this is of interest is that it is a view which seems to be widely shared — though I know of no one else who has articulated it so clearly. Moral scepticism is frequently assumed to be only an option for those who are confident that fully objective descriptions of the world (of some sort, and normally of the sort offered by physics) are possible; to the extent that one becomes sceptical about that, it is felt, one's scepticism about moral values becomes undermined. (For two recent examples, see Snare [7], p. 217, who claims, from a perspective sympathetic to moral scepticism, that 'any reasonable scepticism about morals requires the rejection of radical scepticism about describing and explaining'; and McDowell [3], p. 346, who suggests, from an unsympathetic perspective, that those who are anti-realists about values are motivated by 'a philistine scientism'.)

I wish to question this commonly held view. The issue may be posed as follows. Suppose that initially one is a moral sceptic, in the sense explained, and that one's moral scepticism is held in conjunction with what was referred to as a 'globally unsceptical world-view'. That is, suppose one believes that science, or perhaps some other type of discourse, is in principle capable of describing the world objectively (by which, to repeat, is here meant 'as it is in itself', or 'as it is quite independently of any of our representations of it'), and that moral values are no part of the world as so described. Now, suppose that one's faith in the possibility of such objective descriptions of the world is shaken; is one thereby required to *retract* one's claim that moral values are not part of the world as it is in itself?

The answer is that it depends *in what way* one comes to doubt the possibility of objective descriptions of the world; it depends in what sense this is now thought to be impossible. One might, on the one hand, come to doubt whether the very concept of objectivity presupposed here is coherent – whether it even makes sense to talk of the world 'as it is in itself', as opposed to the world 'as it appears to us' or 'as it is represented by us'. Attacks on this traditional distinction, and especially on the intelligibility of the first term of it, are not uncommon in twentieth-century philosophy. (For some celebrated recent examples, see Davidson [2] and Rorty [5] and [6].) Now, this is evidently one route to the abandonment of the view that descriptions of the world as it is in itself are available, through science or any other method; indeed, one would no longer be able either to accept or reject this view, but would have to regard it as senseless. By the same token, however, one would no longer be able to espouse moral scepticism, nor again to deny it; if there is no coherent notion of the world 'as it is in itself', it makes no sense either to affirm or to deny the claim that moral values are not part of the world as it is in itself. Hence, if the way in which one's 'globally unsceptical world-view' falters is that the concept of objectivity on which it relied no longer seems to one to have sense, one's moral scepticism is indeed bound to falter at the same time and in the same way; for it relies on that same putatively senseless concept.

However, it is by no means clear that this is the *only* way in which one might be led to abandon one's 'globally unsceptical world-view'; it looks as if there is another possibility. One might continue to believe that the *concept* of the world 'as it is in itself' makes sense, but come to doubt whether science, or any other method of discovery, could ever *describe* the world as it is in itself. Several philosophers – notably Williams [8] and Nagel [4] – have recently insisted on the coherence, and the importance, of this notion of an 'absolute' reality; but both Williams and Nagel are understandably cautious about the prospects of attaining the perspectiveless standpoint that would apparently be needed to understand the nature of that absolute reality. Their caution, then, represents a shift from what we have been calling the 'globally unsceptical world-view' – a different kind of shift from that described in the previous paragraph; and indeed, someone might be even more impressed than are Williams and Nagel by the difficulty of transcending the various perspectives which habitually shape our ways of viewing the world, while still maintaining that the notion of the world 'as it is in itself' is perfectly intelligible. Suppose, then, that one's lapse from the initial 'globally unsceptical

cal world-view' takes this form; is one compelled also to jettison one's initial moral scepticism?

It does not seem that one is. Moral scepticism, of the kind we are considering, states that moral values are not part of the world as it is in itself. Now, if one has not rejected the very concept of the world 'as it is in itself', it is not clear why moral scepticism is not just as legitimate an option as it was to begin with. The fact that one has now begun to doubt whether *any* of our ways of talking about the world can hope to capture it as it is in itself seems to do nothing to *undermine* moral scepticism; it merely seems to show that we are as badly off in all areas as we were previously claimed to be in morality.

It appears, therefore, that there is at least one way in which one could continue to be a moral sceptic while ceasing to adhere to a 'globally unsceptical world-view'; and this conclusion runs counter to the line of thought sketched at the beginning. However, the following objection may be raised. Part of that original line of thought was to the effect that, if one gave up one's 'globally unsceptical world-view', moral values would no longer appear as *less real* than other types of entities — but that moral scepticism depended upon this kind of degrading contrast.

This part of the argument, though, appears to trade on a lack of precision. Whether or not moral scepticism depends, as was alleged, on such a contrast itself depends on what exactly the content of moral scepticism is taken to be. One can, indeed, imagine a position which explicitly *referred* to a contrast between morality and, say, science; it might read as follows. 'The entities described by science are features of the world as it is in itself, and moral values are unreal *by comparison with them*'. I see no reason not to regard this as a form of moral scepticism; and this position obviously would be undermined by a weakening of the contrast between morality and science. However, if moral scepticism contains no reference to any areas besides morality, but simply says 'moral values are not part of the world as it is in itself, but are a human product' — which is the way it is more normally portrayed, and the way I have been portraying it here — then it is hard to see why it depends on a contrast between morality and other aspects of the world. Why should it be thought that, in order to claim that moral values are *not* part of the world as it is in itself, one must be able to refer to something *else* which actually *is* the way that one is claiming moral values are not? The idea sounds suspiciously like an instance of the once popular, but surely long discredited, 'paradigm-case' style of argument (I owe this point to Mark Bernstein). In fact, all that seems to be needed is that it should be clear what the moral sceptic is denying — that some kind of account should be available of what it would be like if moral values *were* part of the world as it is in itself. I do not suggest that this is entirely straightforward; but I fail to see why it should require one to adopt an unsceptical any more than a sceptical view about topics other than morality.

One other implication of the original argument, and especially of that part of it I have just been considering, is that, for a view to count as a form of moral *scepticism*, it must impute to morality some *fault* or *shortcoming*; and this imnplication I am happy to accept. However, this is quite consistent with my position. If morality's being merely a human product is a shortcoming of morality, then that is so regardless of the status of science. If it matters to us that moral values should not be a human product, then morality is in bad shape if it *is* a human product. If it turns out that the status of science is no different from that of morality, that does not show that morality is *not* in bad shape after all; if anything, it simply shows that science is in bad shape as well. Again, the moral sceptic will need to be able to explain clearly what it is that morality is lacking; but it is not clear why this would require pointing to something else that is *not* lacking in the relevant respects.

To sum up: Suppose that one's initial position can be represented schematically as

(1) Science (or some other type of discourse) is capable of describing the world as it is in itself, and
(2) Moral values are not part of the world as it is in itself.

Suppose then that one abandons (1), and instead adopts

(1′) While the notion of the world 'as it is in itself' is an intelligible one, it is doubtful whether we can ever hope to describe the world as it is in itself, either through science or any other method.

What I have been arguing is that there is no reason why (2) should not be held in conjunction with (1′), just as much as with (1), nor any reason why (2) should not count as a form of moral scepticism when held in conjunction with (1′), just as much as when held in conjunction with (1). If I am right, it is not the case, as stated by Annas and implied by others, that (modern) moral scepticism is essentially local. Undeniably, moral sceptics often are unsceptical about matters other than morality, and especially about the capabilities of science (and it would be interesting to examine the reasons for this), but this is not something that is *required* of them qua moral sceptics.

I should emphasize, in closing, that this conclusion applies only to the specific form of moral scepticism discussed here. This is probably the view most often referred to as 'moral scepticism' in contemporary moral philosophy, but it is by no means the only view to receive that label. A fuller discussion would certainly need to consider, among other things, the precise significance of the term 'moral scepticism' itself, and indeed, of the term 'scepticism' in general; for this will affect which views it is legitimate to describe as forms of moral scepticism – a matter which, it seems to me, is not well understood – which in turn will affect whether or not my conclusion can be generalized.[1]

University of Texas at Arlington,
P.O. Box 19527,
Arlington, TX 76019, U.S.A.

© RICHARD BETT 1988

REFERENCES

[1] Julia Annas, 'Doing Without Objective Values: Ancient and Modern Strategies', in *The Norms of Nature*, ed. Malcolm Schofield and Gisela Striker, Cambridge University Press, 1986.
[2] Donald Davidson, 'On the Very Idea of a Conceptual Scheme', *Proceedings of the American Philosophical Association* 47 (1973–4), 5–20.
[3] John McDowell, 'Virtue and Reason', *The Monist* 62 (1979), 331–50.
[4] Thomas Nagel, *The View from Nowhere*, Oxford University Press, 1986.
[5] Richard Rorty, *Philosophy and the Mirror of Nature*, Princeton University Press, 1979.
[6] Richard Rorty, *Consequences of Pragmatism*, University of Minnesota Press, 1982.
[7] Francis Snare, 'The Empirical Bases of Moral Scepticism', *American Philosophical Quarterly* 21 (1984), 215–25.
[8] Bernard Williams, *Descartes: The Project of Pure Enquiry*, Penguin Books, 1978.

---

[1] I would like to thank Julia Annas for her helpful comments on a previous version of this paper.

SOME BY THE WAY REMARKS ON WREEN'S 'BY' WAYS

By KARL PFEIFER

MICHAEL Wreen ('Two "By" Ways', ANALYSIS 47.2, March 1987, pp. 120–4) worries about causal loops apparently entailed by some action sentences containing 'by' locutions. These 'paradoxical' loops can be avoided, he contends, by noting that 'by' may indicate either (i) being caused or generated by causal or other sorts of Goldmanian action generation, or (ii) being indirectly achieved or performed by.

This is a specious solution. On the one hand, the kind of loop Wreen has drawn our attention to is inherent in 'by'-locutionary talk generally and not just in the particular kind of causal case he has addressed. On the other hand, the very distinction Wreen seeks to draw is subverted by the fact that indirect performance and achievement are themselves instances of generation on Goldman's account (expounded in *A Theory of Human Action*, Princeton University Press, 1970; hereafter '*THA*'). And we shall see that Goldman makes a poor ally in any case.

Let me elucidate the problem with a bare bones example. Consider a game, ritual, or whatnot involving the movement of counters of various colours and materials. John moves a black wooden counter, rendering both the following true:

(1) John moves a wooden counter by moving a black counter.
(2) John moves a black counter by moving a wooden counter.

Since the black counter moved is the wooden counter moved, the difference between (1) and (2) seems clearly to be a pragmatic matter of emphasis relative to one's explanatory concerns. And because one can have more than one such concern, one can have it both ways. So the 'by' relation here can be understood as an explanatory relation, sensitive to wording and context, sans paradox. Furthermore the loop in this example can easily be made a matter of indirect performance or achievement.

Suppose John wants to move a black piece because he feels black is lucky. He knows he has to move a wooden counter to do so, since all the black counters of plastic or rubber are gone. To move something black, he must 'overshoot' and move something black-and-wooden. (John is like Wreen's character, Walt, who proves that all natural numbers have property P by overshooting and proving that all real numbers have property P.) Now suppose further that Bill, another participant in this 'counter activity', views the movement of a wooden counter as unlucky and that John recognizes that by moving something black he will perform the from Bill's point of view unlucky act of moving a wooden counter (this would be like Walt's proof, its overshooting aforethought notwithstanding, also amounting to a proof that the naturals have P only if the reals do). Hence we now have a simple indirect-performance-or-achievement loop.

Is there anything wrong with such a loop? No, if one takes the 'by' relation to be explanatory and explainable in ways unfettered by generational presumptions. Yes, if this 'by' relation is construed as Goldman's action generation. That technical transmogrification of a commonplace idiom leads to serious trouble.

Generation is stipulated to be asymmetric, transitive, and irreflexive (*THA*, p. 21). Moreover it only holds among acts or 'act-tokens' construed and finely individuated as property-exemplifications by an agent at a time (*THA*, p. 10). Yet (1) and (2) above each satisfy Goldman's criterion for the occurrence of the kind of generation he calls 'simple generation'. 'In simple generation', says Goldman, 'the existence of certain circumstances, conjoined with the performance of *A*, ensures that the agent has performed *A'*.' (*THA*, p. 26) The relevant circumstances in this case would include the coincidence of blackness and woodenness in the same counter; this in itself is sufficient to give us the mutual generation of John's act of moving a wooden counter and John's act of moving a black counter. Circumstances for simple generation can also subsume intentions, motives, and beliefs (*THA*, p. 27), the sorts of things that go into characterizations of 'action plans' (Wreen's phrase). So indirect performance or achievement is in the current context just a special case of simple generation. Such mutual generation — a generational loop — either violates the asymmetry of generation, or, if one countenances the generation of two distinct act-tokens involving the same act-property *et seq.*, violates the identity conditions for act-tokens. Hence examples such as the present one show the generational account to be fundamentally incoherent.[1]

If such a malign loop is to be avoided we must either deny the 'by' relation the offending generational properties, or, failing that, we must give up the idea that particular acts can be straightforwardly identified with exemplifications of act-properties as ascribed by the verb phrases of ordinary action sentences. But if we do that, it is less clear how there was a causal loop problem to begin with.

University of Saskatchewan,
Saskatoon, Canada S7N 0W0

© KARL PFEIFER 1988

---

[1] There are no easy amendments for avoiding this, either. For detailed discussion of this problem see my 'A Problem of Motivation for Multipliers', *Southern Journal of Philosophy*, 20 (1982) 209–24, and 'A Consideration of Modifications to the Multiplying Account', *Philosophy Research Archives*, 11 (1985) 141–54.

A REPLY TO CAHN

By DANIEL H. COHEN

STEVEN M. CAHN (ANALYSIS 47.3, June 1987, pp. 175–6) asks how a principled divestiture of stocks is morally possible. The problem, correctly noted, derives from two premises. First, ownership of stock in a company whose policies are morally objectionable is itself morally objectionable. Second, it cannot be morally obligatory to perform an act which requires someone else to commit an offence of identical gravity. Because stocks can only be sold when there is a buyer, selling any such stocks accomplishes no net reduction of morally objectionable economic behaviour in the world. Divestiture, the argument concludes, is something of an ostrich act. It redistributes guilt instead of eliminating it. There can be no obligation to sell these stocks. It might be added that there are other, better models for moral behaviour: even if it is supererogatory to take the sins of the world upon oneself, we can at least be expected to refrain from casting our sins onto the shoulders of others.

For all its validity, the argument is specious, a straw man argument. The purpose of calls for divestiture of stocks in certain companies is not to cleanse individuals' portfolios and souls. The purpose is to bring pressure to bear on the companies in question to alter the odious practices. The line of reasoning is that a general increase in the numbers selling a given stock will increase the supply of that stock readily available on the market and so have a downward effect on the price of the stock. This would be the kind of message readable by even the most completely economic of agents. Divestiture is morally possible because it is a means, not an end. Accordingly, divestiture is successful only if it succeeds in having an effect. The divestiture that Cahn seems to have in mind would be successful as soon as the agent in question no longer owns the stock. His assessment of this is right: there is only a dubious moral value to be found in that. Influencing the policies of an Immoral-Corp can be a righteous goal; a clean portfolio is merely a self-righteous one. Of course, the two goals call for the same strategy, but identical strategies do not necessarily indicate the same goal. The distinction is not a new one: divestiture merely to clean one's own hands may be compatible with duty but no more.

Colby College,
Waterville, Maine 04901, U.S.A.

© DANIEL H. COHEN 1988

RAMSEY-LEWIS IS BETTER THAN MACKIE

By DAVID PAPINEAU

IN a recent article, 'Laws and Accidents' [1], I had occasion to compare two different Humean accounts of the law-accident distinction, the first due to J. L. Mackie [2], and the second originally formulated by F. P. Ramsey [3], and more recently revived by David Lewis [4]. In that article I suggested that these two accounts are more or less equivalent. But further reflection on the comparison has persuaded me that the Ramsey-Lewis view is preferable to Mackie's.

Ramsey said that laws are 'consequences of those propositions which we should take as axioms if we knew everything and organized it as simply as possible in a deductive system' ([3], p. 130). We can spell out his thought as follows. One possible way of getting all the general truths into a deductive system would be to take them all as axioms. But that wouldn't be simple. On the other hand, simpler systems, with less axioms, would tend to leave some general truths dangling, in the sense that those truths would be neither axioms nor theorems of the simpler system. The optimal system would be the system which best combined simplicity with the inclusion of as many general truths as possible. The Ramsey–Lewis view is that laws are those general truths that appear as axioms or theorems in the optimal system, while accidents are those general truths that are left dangling by that system.

Mackie's view is that laws are differentiated from accidents by being inductively supported by their instances. The predicates in laws are mutually projectible, but those in accidents are not.

In 'Laws and Accidents' I tried to bring these two accounts closer together by adopting the assumption that projectible predicates are those that can be defined in terms of predicates that appear in the axioms of the optimal system. The suggested motivation for this assumption was that (a) only generalizations involving such predicates can be theorems in the optimal system, and (b) it seems sensible to expect an observed pattern to continue just in case it might be explicable by fundamental scientific truths.

Given this assumption, Ramsey–Lewis lawhood implies Mackie lawhood, since predicates in theorems of the optimal system will always be projectible. Mackie lawhood doesn't quite imply Ramsey–Lewis lawhood, however, since a general truth may be framed in predicates which are definable in terms of the basic predicates of the optimal system, without actually being a theorem of that system.

In the earlier paper I suggested that Mackie laws which weren't Ramsey–Lewis laws would be the exception rather than the rule, in that the vast majority of true, projectible generalizations could be expected also to be theorems of the optimal system. And having related the two accounts to this extent, I left it at that. But it now strikes me that if we concentrate on the exceptional cases — that is, on the true projectible generalizations which *aren't* theorems of the optimal system, and which thus differentiate the Ramsey–Lewis and Mackie views — then a strong case can be made in favour of the Ramsey–Lewis view.

Imagine that Dr Jane Jones is conducting an experiment involving some electrical apparatus, and that she notices that, when the apparatus is switched on, a distinctive kind of periodic signal is received on the medium wave radio she is listening to. She doesn't have an explanation for this, but naturally enough assumes it to be a lawlike, counterfactual-supporting fact.

But now suppose that the general claim, 'Whenever this kind of apparatus is switched on, there are radio waves of a distinctive pattern in the vicinity', isn't a law, but just an accident. Imagine, for example, that there isn't any causal connection between the experimental apparatus and the radio signal: the radio signal is due to something quite different, which just happens to be present whenever the apparatus is on. (If we like, we can beef up the example by imagining that Dr Jones's experiment isn't the only case in history where an apparatus of that kind is used, but that in the other cases too it so happens that there is the radio signal whenever the apparatus is on.)

In the circumstances as described, the generalization won't be a Lewis law, since it won't follow from any of the fundamental principles of physics. But it will be a Mackie law, for it will be true, and it will be framed in projectible terms — the terminology of electrical circuits and radio waves is projectible, if anything is. In fact, of course, it's not a law. So it's a counter-example to the Mackie view.

No doubt Dr Jones, in taking it the generalization to be a law, will *believe* it to be derivable from fundamental physics. But that's beside the point. The generalization *isn't* derivable from fundamental physics, and so Dr Jones is mistaken about this, just as she is mistaken about the generalization being a law. She wouldn't be mistaken to believe that the generalization was a *Mackie* law, of course: for it is true, and framed in projectible terms. But presumably that's not what she believes, in believing the generalization to be a law. Which illustrates, once more, how the Mackie account fails to capture the notion of a real law.

Dept. of History and Philosophy
of Science,
Free School Lane,
Cambridge CB2 3RH

© DAVID PAPINEAU 1988

REFERENCES

[1] David Papineau, 'Laws and Accidents', in G. Macdonald and C. Wright, eds, *Fact, Science and Morality*, Basil Blackwell, 1986.
[2] J. L. Mackie, *Truth, Probability and Paradox*, Clarendon Press, 1973.
[3] F. P. Ramsey, 'General Propositions and Causality', 1929, in his *Foundations*, edited by D. H. Mellor, Routledge and Kegan Paul, 1978.
[4] David Lewis, *Counterfactuals*, Basil Blackwell, 1973.

RATIO

Ratio will no longer appear in exactly its previous form. In its place Basil Backwell will publish from June 1988 a journal as continuous as possible with the former **Ratio.** The journal will publish high quality work on a wide variety of topics. The attempt to further co-operation and understanding between philosophers who work in German and philosophers who work in English will remain an important feature of editorial policy.

CALL FOR PAPERS

Submission of papers for publication in this journal is now invited. Typescripts for the editor's consideration should be sent to:
Edward Craig, Churchill College, Cambridge CB3 0DS, England

The editor welcomes submissions in German or in English. German submissions, if accepted, will be translated into English at the expense of the publisher. Contributors should retain a copy of work submitted; the editor regrets that papers cannot be returned.

Basil Blackwell · Oxford and New York

Bioethics

CALL FOR PAPERS

Special Issue for 1989

The Scientific, Philosophical and Ethical Uncertainties of Pain

There is a range of issues relating to pain which cross the boundaries of science, philosophy of mind and ethics, and have implications for our treatment of foetuses, infants, older humans in abnormal conditions, and non-human animals. What can we know about whether, and to what degree, beings in these categories can feel pain? And what are the ethical implications of any uncertainty we may have about the pain they feel?

We invite contributions on any aspect of these issues, or on related ethical problems in biomedical research or medical or veterinary practice. Such problems include, but are not limited to, animal experimentation, medical uses of the human foetus, late abortions, and painful procedures on newborn infants and on older patients who are unable to consent. Contributors who are in doubt about whether their topic will be suitable for the issue are invited to send us an outline of their work.

The deadline for submission of papers is January 15, 1989.

Helga Kuhse and Peter Singer, Centre for Human Bioethics, Monash University, Clayton, Victoria 3168, Australia.

Basil Blackwell · Oxford and New York

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, Jane Heal, R. F. Holland, C. Hookway, C. A. Kirwan, D. H. Mellor, H. W. Noonan, M. A. Stewart, A. R. White, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices for 1988 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£13.00	£16.00	U.S. $33.00
Individuals	£8.00	£10.00	U.S. $21.00

Single Issue Prices

	U.K.	Overseas	U.S.A. & Canada
Institutions	£4.30	£5.10	$10.60
Individuals	£3.10	£3.70	$6.70

5 APR 1988

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

CONTRIBUTIONS. Articles submitted for publication should be addressed to the Editor, Mr Peter Smith, Department of Philosophy, University College of Wales, Aberystwyth, Dyfed SY23 3DY.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

Vol. 48 No. 3 (New Series No. 219) June 198

ANALYSIS

Edited by
Peter Smith

CONTENTS

Playing games with eternity: the Devil's offer Edward J. Gracel
Indicative and subjunctive V. H. Dudma
Modus Tollens revisited Ernest W. Adam
Vague identity: Evans misunderstood David Lewi
Vagueness and identity B. J. Garret
A note on 'Most' John Slane
The chaology of mind Adam Morto
The Waterfall Illusion Tim Cran
Crane's Waterfall Illusion D. H. Mello
Concepts in perception Tim Cran
The 'by' word Michael Wree
On a theory about theories about all theories George Re
I'm glad you asked me that question Lawrence Zalcma

BASIL BLACKWELL OXFORD AND NEW YORK

ISSN 0003-2638

BLACKWELL

Moral Vision

An Introduction to Metaethics

DAVID McNAUGHTON

This textbook introduces students to many of the major issues in ethics, by way of a clear exposition of both sides of the contemporary debate between moral realists and non-cognitivists.
288 pages, hardback **£25.00** (0 631 15408 6)
paperback **£8.95** (0 631 15945 2)

Moral Dilemmas

WALTER SINNOTT-ARMSTRONG

In the first full-length discussion of moral dilemmas, Professor Sinnott-Armstrong argues that unresolvable moral dilemmas are both possible and actual.
Philosophical Theory Series
240 pages, hardback **£27.50** (0 631 15708 5)

Weakness of Will

An Introduction

WILLIAM CHARLTON

William Charlton provides a comprehensive, up-to-date introduction to both ancient and contemporary discussions of irrational behaviour, and offers his own teleological view of human conduct as explainable in terms of reasons and purposes.
208 pages, hardback **£25.00** (0 631 15758 1)
paperback **£8.95** (0 631 15759 X)

Contemporary Social Philosophy

GORDON GRAHAM

This book introduces students of philosophy and the social sciences to the wide range of issues, both theoretical and practical, which occupy contemporary social philosophers.
224 pages, hardback **£25.00** (0 631 15705 0)
paperback **£7.95** (0 631 15986 4)

Ideas of Human Nature

An Historical Introduction

ROGER TRIGG

'Professor Trigg's book is clear, accurate and absorbing, well worth recommending to A-level students, undergraduates and anyone seriously interested in finding out what sort of thing we might be thought to be.'
Professor Stephen Clark, University of Liverpool
208 pages, hardback **£25.00** (0 631 14533 8)
paperback **£7.95** (0 631 14534 6)

Basil Blackwell
108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016.
Toll-free ordering: 1-800-638-3030 (USA)

PLAYING GAMES WITH ETERNITY: THE DEVIL'S OFFER

By EDWARD J. GRACELY

SUPPOSE Ms C dies and goes to hell, or to a place that seems like hell. The devil approaches and offers to play a game of chance. If she wins, she can go to heaven. If she loses, she will stay in hell forever; there is no second chance to play the game. If Ms C plays today, she has a 1/2 chance of winning. Tomorrow the probability will be 2/3. Then 3/4, 4/5, 5/6, etc., with no end to the series. Thus every passing day increases her chances of winning. At what point should she play the game?

The answer is not obvious: after any given number of days spent waiting, it will still be possible to improve her chances by waiting yet another day. And any increase in the probability of winning a game with infinite stakes has an infinite utility. For example, if she waits a year, her probability of winning the game would be approximately .997268; if she waits one more day, the probability would increase to .997275, a difference of only .000007. Yet, even .000007 multiplied by infinity is infinite.

On the other hand, it seems reasonable to suppose the cost of delaying for a day to be finite — a day's more suffering in hell. So the infinite expected benefit from a delay will always exceed the cost.

This logic might suggest that Ms C should wait forever, but clearly such a strategy would be self defeating: why should she stay forever in a place in order to increase her chances of leaving it? So the question remains: what should Ms C do?[1]

© EDWARD J. GRACELY 1988

Department of Community and Preventive Medicine,
The Medical College of Pennsylvania,
3300 Henry Avenue,
Philadelphia, PA 19129, U.S.A.

---

[1] I would like to thank Janet Fleetwood for her very helpful comments on the first version of this paper.

INDICATIVE AND SUBJUNCTIVE

By V. H. DUDMAN

THE idea that English has more than one declarative "mood" has been dismissed as superstitious by empirically-minded grammarians of English for centuries[1] — with such spectacular unsuccess, however, that the indicative/subjunctive dichotomy stands today as a cornerstone for philosophical and logical speculation about "conditionals". Let me be next into the breach. I shall urge that there is no grammatical basis for any such distinction. And as for the particular adjudications of mood logicians and philosophers actually propose, there is neither rhyme nor reason to them (Parts I and III). My bent, then, is basically destructive. But I shall also be outlining a better taxonomy (Part II).

I

We can begin by observing that each of (A) and (B) can convey (at least) two different *messages*:

(A) If she misses the last bus the Countess will walk home

(B) If she missed the last bus the Countess would walk home.

(A) can convey either a generalization m_1 about the present (cf. '. . . she will often walk home') or a more particular claim m_3 about the future (cf. 'If she misses the last bus this evening . . .'). (B) can convey a generalization m_2 which is just like m_1 except for holding in the past, or it can convey a claim m_4 about the future which is only subtly different from m_3.

Now, philosophers and logicians diagnose the indicative in m_2's case and the subjunctive in m_4's, whereupon mood cannot possibly be a trait of sentences, since the sentence (B) is the same said either way. It must be the messages that are either indicative or subjunctive then; and indeed, according to the received view, m_1, m_2 and m_3 are all indicative and m_4 subjunctive.

But what a curious distribution! Why should m_4 count as different from the other three? What can be so special about m_4 that makes m_3 fundamentally more like m_1 and m_2? Nobody says.

The decision is the more bewildering seeing that m_3 fits into a pattern into which m_4 also fits. The pattern emerges when we compare m_3 and m_4 with the future interpretation m_5 of (C), the present interpretation m_6 of (D), the present interpretation m_7 of (E) and the past interpretation m_8 of (E):

---

[1] Concerning perceptions of the subjunctive and cognate moods down the years, see I. Michael, *English Grammatical Categories and the Tradition to 1800*, Cambridge: Cambridge University Press, 1970, 424–35. Apparently first blood goes to James Greenwood (d. 1737), Sur-master at St. Paul's School London, who in his treatise *An Essay towards a Practical English Grammar (describing the genius and nature of the English tongue. Giving likewise a rational and plain account of grammar in general, with a familiar explanation of its terms)*, London, 1711, urged: 'In English there are no Moods, because the verb has no diversity of Endings'. Greenwood's aspirations, formulated in his parenthetical advertisement, awaken stirring chords across the centuries.

(C) If the auditors had come tomorrow they would have found our books in order

(D) If the Duke was here he would be furious

(E) If the Duke had been here he would have been furious.

The pattern is that the time, future present or past, that a clause or sentence is about is always *later* than the time, present past or past past, registered by the form of its predicate. Thus m_3 relates to the future while the 'will' and 'misses' of (A) register presentness; m_4 relates to the future, while the 'would' and 'missed' of (B) register pastness; m_5 again relates to the future, while in (C) 'would have' and 'had come' register past-pastness (pastness with respect to an already past moment, perhaps the moment when the auditors decided to come early). And so on. This description surely demarcates a compelling natural kind, yet received adjudications of mood slice clean through it.

The mystery deepens yet further when we consider (F), whose natural interpretation m_9 is utterly unlike any of m_1 to m_8:

(F) If Socrates is a man, Socrates is mortal.

Under m_9, the strings 'Socrates is a man' and 'Socrates is mortal' are each understood exactly as if encountered standing alone between full stops. Under none of m_1 to m_8 is there a string in the scope of 'if' which is understood as if standing alone. This novelty notwithstanding, m_9 is obliged to submit to the discipline of mood — and finds itself assigned to the indicative.

The natural interpretation m_{10} of (G) also awaits arbitration:

(G) If Tom was fat his sister was immense.

It is different from all the foregoing and indeed sui generis. There are again two prior sentences, m_{10} in this regard like m_9. But unlike m_9, m_{10} insists on a particular order: one cannot venture it by saying

Tom's sister was immense if Tom was fat.

In this regard (G) differs from every comparable English locution I can think of.[2] Is m_{10} to be reckoned indicative or subjunctive? Not subjunctive, presumably. So unless we want a third mood it too must be indicative, for certainly (G) occurs declaratively under m_{10}.

Now *who*, presented with the task of classifying m_1 to m_{10} grammatically, would respond by cutting m_4 to m_8 out of the herd, and taking the division between them and the rest as *the* fundamental dichotomy? What conceivably could be the essential grammatical distinction producing this particular scatter? "The rest" are m_1, m_2, m_3, m_9 and m_{10}, and what could possibly be common to them which was more compelling than the tense-time pattern of m_3 to m_8 which this analysis sunders? What, indeed, is common to m_1 and m_2, on the one hand, and m_9 on the other? Or to m_3 and m_9?

---

[2] Before 1986 I wrongly classified m_{10} with m_9, solely on the basis of the two prior sentences.

To my best awareness none of these questions has ever been tackled, much less answered. It is perhaps the worst aspect of current linguistic speculation among logicians and philosophers that it relies crucially upon a handful of technical terms which are *nowhere defined or explained.* 'Indicative' and 'subjunctive' still await grammatical explications from logicians and philosophers, who habitually launch straight into semantics, seeming not to appreciate that they thereby leave their theories without content, for the simple reason that until we have been told what 'indicative' means in non-semantic terms, we shall not even know what So-and-so's semantic theory of "indicative conditionals" is *about.*[3]

II

What we casually call 'if'-sentences have 'if' occurring in them as a "subordinating conjunction".[4] Now, a word or phrase can occur as a subordinating conjunction in two utterly different ways: either *externally prefixed to a sentence* or as *first word of a subordinate clause*:

(H) As/(al)though/because/provided that/since/unless/whereas Socrates is a man, Socrates is mortal

(I) After/as soon as/immediately/once/provided that/the minute/unless/until/when(ever) she misses the last bus, the Countess walks home.

As the reader has doubtless anticipated, 'if' is in this regard versatile. Under m_9, (F) has 'if' externally prefixed to the sentence 'Socrates is a man'. But under its natural present-habit interpretation m_{11}, (J) is a *subject-predicate* sentence with a subordinate clause 'if she misses the last bus' embedded in its predicate:

(J) If she misses the last bus the Countess walks home.

The latter perception is reached as follows. Whereas

> The Countess [cf. usually/often/occasionally/rarely/never] walks home

alleges the Countess's present propensity to walk home, (J) alleges her present propensity to *walk home if she misses the last bus.* 'If she misses the last bus' in (J) is like 'On Wednesdays' in

---

[3] Most leading exponents of 'if' among contemporary logicians and philosophers accept the indicative-subjunctive distinction as basic without attempting to say what it consists in. To name names here would be invidious.

[4] See e.g. R. Quirk *et al., A Grammar of Contemporary English,* London: Longman, 1972, 720, 727ff, 743ff. Cf. R. D. Huddleston, *Introduction to the Grammar of English,* Cambridge: Cambridge University Press, 1984, 339 and passim.

On Wednesdays the Countess walks home.

And what conditions the choice of conjunction, 'if' rather than 'after' or 'as soon as', is the particular refinement of habit whose present currency is alleged. Under m_{11}, then, (J)'s immediate constituents are a subject, 'the Countess', and a predicate comprising everything else. This predicate is *complex*, because it contains not only a verb phrase 'WALK home' but also an entire 'if'-clause.

But (F)'s immediate constituents under m_9 are not a subject and a predicate. They are the *independent sentence* 'Socrates is mortal' and the string 'if Socrates is a man', the latter in its turn obtained by externally prefixing 'if' to a thereupon *dependent sentence* 'Socrates is a man'. Under m_9, (F) is parsed as a *compound sentence*, a sentence compounded using discrete prior sentences.

Now, there can be no greater grammatical difference than that between a message encoded as a subject-predicate sentence and a message encoded as a compound sentence. Why, they disagree even in outermost structure. *This*, not indicative/subjunctive, is the fundamental grammatical dichotomy for interpretations of 'if'-sentences.

In my submission, English generates what we casually call 'if'-sentences in just four basically different ways, and accordingly the messages which English encodes as 'if'-sentences divide into four grammatical *categories*.[5] All four are represented by m_1 to m_{11}: the first by m_9, the second by m_{11}, m_1 and m_2, the third by m_3 to m_8, and the fourth by m_{10}.

As regards the fundamental grammatical dichotomy just proclaimed, we have already seen that the first and fourth categories belong on one side, and the second on the other. But what of the third? Under m_3, is (A) a compound sentence or a subject-predicate one?

The third category, I should add, uses all the modals, not just WILL. It embraces future interpretations of sentences of the form

> If *a U*-s . . . , *b* will/can/may/shall/must/should$_1$/ought to/needn't/daren't *V* . . . ,

present and future interpretations of sentences of the form

> If *a U*-ed . . . , *b* would/could/might/should$_2$ *V* . . . ,

and past, present and future interpretations of the form

> If *a* had *U*-en . . . , *b* would/could/might/should$_2$ have *V*-en . . . ,

together with other messages involving slightly more recondite temporal arrangements.

Now, each of the several sentences (K) has a future interpretation:

---

[5] When I recognized only three, I was conflating the first and the fourth. See e.g. 'Parsing *if*-sentences', *Analysis* 44 (1984) 145–53, esp. 147f.

(K) The Countess will/can/may/shall/must/should$_1$/ought to/needn't/daren't walk home.

These future messages are all *judgments* (let me say) concerning the Countess's walking home at a time later than that registered by the choice of inflectional form at the start of the predicate, and it is the *kind* of judgment that conditions the choice of modal, WILL, CAN, MUST or whatever. Now, the effect of introducing the subordinate clause 'if she misses the last bus' into (K) is to modify the judgment:

The Countess will-if-she-misses-the-last-bus walk home.

The judgment, still about the Countess's walking home at a time later than that registered by the form, has merely become more recondite.

Under m_3, then, (A)'s immediate constituents are a *subject*, 'the Countess', and a *predicate* comprising everything else. This predicate is complex, for besides containing both a verb phrase and an inflectional form of a modal, it contains an entire 'if'-clause. (The basic grammatical difference between the second category and the third is that in the second the subordinate clause governs the verb phrase whereas in the third it governs the modal.) This answer will come as a hard saying to philosophers and logicians, who ensure at a stroke the grammatical untenability of their semantic theories by diagnosing "antecedents" and "consequents" in the third category.

Of course, the key problem posed by the third category is to explain the temporal relation which is its hallmark: what role in the third-category message is played by the time registered by the form? Something for another day.[6]

III

According to my four-way analysis in Part II, the "subjunctive" is merely a proper subset of the third category. Even *within* the third category, the messages called 'subjunctive' are merely the past-tense ones, and no more special or different, according to me, than past-tense messages in general. And who would have thought that mere tense might be an arbiter of mood? But this fact has gone unnoticed because the very idea of the subjunctive clouds perception of tense: 'subjunctive' encourages us to treat the forms of predicates as registering something special, different, *not* tense. The tense-time pattern that unifies the third category has effectively been hidden from logical and philosophical gaze by the mystical power of a word.

---

[6] I confront this problem in 'Conditional interpretations of *if*-sentences', *Australian Journal of Linguistics* 4 (1984) 143-204, and again in 'Towards a theory of predication for English', *AJL* 5 (1985) 143-96, sections 34-7. I explore first-category messages in 'Antecedents and consequents', *Theoria* 52 (1986) 168-99, especially in Part 4.

Perhaps what sustains philosophers and logicians in their adjudications of mood is an inchoate but reassuring "Latin Prose Theory":[7] just those conditionals are indicative (subjunctive) which translate into indicative (subjunctive) conditionals of Latin. Of course, the fundamental objection to the Latin Prose Theory is simply that it is a crazy way of coming at English: when busy with the grammar of *one* language, why take on a whole new set of empirical problems presented by the grammar of *another* language, *plus* all the imponderables of faithful translation between the two? Actually, since the matter has come up, it seems that translation between English and Latin is none too faithful. Compare, for example, m_8 with the natural interpretation m_L of its customary "pluperfect subjunctive" rendition into Latin. Received classical doctrine has it that m_L "rejects" the proposition that the Duke was present: m_L is in this sense "counterfactual".[8] But nothing like this applies to m_8, which can be affirmed in circumstances where the Duke's absence is by no means acknowledged.[9] It makes perfect sense to say

> Whether or not the Duke actually *was* here, if he *had* been here . . .

If the earnests of our classicist colleagues are to be trusted, therefore, m_8 and m_L differ in a quite elemental way, products, evidently, of quite different systems of thought. – Still, this is an aside: the prior objection to the Latin Prose Theory of English mood is the centuries-old one: that there is absolutely no warrant for imposing current grammatical doctrine about one language upon grammatical speculation about another.

In their assimilation of m_9 and m_{11} as both "indicative", logicians and philosophers are perhaps encouraged by the presentiment that second-category messages are quantificational elaborations, somehow, of first-category ones – maybe involving quantification over temporal instants. But quantificational analyses of the second category are unworkable. They will not extend to cover past and past past habits, or to cover the congruence of m_{11} with the various natural interpretations of (I), *or* to allow for adverbs of frequency, as in

> If she misses the last bus she always/customarily/normally/usually/occasionally/seldom/rarely/never walks home.

---

[7] The idea for this paragraph, and the term, come from the Editor, to whom I am grateful for other help as well.

[8] E.g. B. L. Goldersleeve and G. Lodge, *Latin Grammar* 3rd edition, New York: Macmillan, 1895; reprinted 1971; see 84ff.

[9] This fact about the English language is by now well advertised among logicians and philosophers. See e.g. R. M. Chisholm, 'The contrary-to-fact conditional', in M. Feigl and W. Sellars (eds), *Readings in Philosophical Analysis*, New York: Appleton Century Crofts, 1949, 482–97, esp. 483. See also A. R. Anderson, 'A note on subjunctive and counterfactual conditionals', *Analysis* 11 (1951) 35–8.

And in any case, they get the outermost structure wrong.

In the eyes of logicians and philosophers themselves, the crucial assimilation is that of m_3 and m_9 as both indicative, and it seems that sometimes they are encouraged in this by the observation that those prepared to assert (L) when knocked unconscious are found maintaining (M) when they come round:

(L) If Oswald does not shoot Kennedy someone else will

(M) If Oswald did not shoot Kennedy someone else did.

Inferring that (L) is just a future version of (M), they hazard the same analysis for both. The reasoning is less than compelling, since those prepared to assert (L) before the event are also prepared to assert (*N*) now, provided they have not changed their minds:

(N) If Oswald had not shot Kennedy someone else would have.

And surely the true parallel is between (L) and (N): the proponent of (N) is only saying what could have been said as (L) at some earlier time. (M) is the odd one out: the mere fact that someone shot Kennedy is a sufficient ground for asserting (M), and thus (M) is found assented to even by those who withhold credence from the conspiracy alleged in (L) and (N).[10]

Incidentally, according to logicians and philosophers, (A) ought to read 'If she *will* miss the last bus . . .' to express m_3, but evidently no one sees any problem, since those who notice the discrepancy casually dismiss it as a perverse idiosyncracy *of English*. English, it seems, says 'misses' in (A) when what it really means is 'will miss'.[11] What the survival value of this irregularity could possibly be is not conjectured. In point of fact there is no truth in this ad hoc excuse, witness the difference between saying 'If we miss the last bus (we will have to walk)' and saying 'If we will miss the last bus (there's no point in our running)'.

In present-day English (O) is a *stylistic variant* of (D) under m_6:

(O) If the Duke were here he would be furious.

Quite generally across the third category, 'were' can freely be substituted for 'was' in the 'if'-clause.[12] Not vice versa though.

Logicians and philosophers perhaps need to be warned that some grammarians apply the term 'subjunctive' directly to singular 'were'.[13] According to their usage, we have the subjunctive in (O) but the *indicative* in (D). Unsurprisingly, these thinkers hold the subjunctive 'unimportant': with the natural interpretations of (D) and (O) identically m_6, this "subjunctive" is of no consequence whatever.

---

[10] Compare T. J. Smiley, 'Hunter on conditionals', *Proceedings of the Aristotelian Society* 84 (1983) 169–77, especially 176.

[11] Thus H. Reichenbach, *Elements of Symbolic Logic*, New York: Macmillan, 1947. See 288: *will miss* 'would be more correct'.

[12] Compare O. Jespersen, *A Modern English Grammar on Historical Principles Vol. IV*, London: George Allen & Unwin, 1931, 129–39.

[13] Thus for example Quirk *et al.*, *op. cit.*, 409.

Of course, logicians and philosophers diagnose the subjunctive in (D) and (O) alike. Nevertheless, many seem to regard singular 'were' as a "sign" of their subjunctive. This is demonstrably wrong, for singular 'were' is also an option in cases where logicians and philosophers recognize *no* subjunctive:

> He had been an ass to suppose Captain Steerforth a murderer; it was part of the confusion of mind caused by Dr Drake's pills. And if Captain Steerforth were innocent, then he was a potential, a natural ally against his enemies [Evelyn Waugh, 1957].
>
> If he happened to be in the midtown area at noon, it was a convenient place for lunch, and, if Annabel were playing bridge late, a pleasant one for a solitary evening drink [Louis Auchincloss, 1973].
>
> (P) We discussed what was best to be done in our circumstances. Violet put forward a proposal: if Parliament were recalled we would set off for home at once [Anthony Powell, 1980].

(Logicians and philosophers ought not to diagnose the subjunctive in (P), because its 'if'-sentence is merely a "backshifted" version of 'If parliament is recalled we will etc.') In short, there is no firm connection between the entirely optional device of singular 'were' and what logicians and philosophers want to call the subjunctive.

There is a scattering of rather specialized devices each of which deserves mention only because some grammarian has styled it 'subjunctive'.[14] For instance, there is the "base subjunctive", so called because the base of some verb occurs where a finite form might be expected:

> (Q) If musik *be* the food of love . . .
>
> (R) It is necessary that every passenger *attend* boat drill.

I recognize two varieties here. (Q) is merely a stylistic variant of 'If musik is the food of love, . . .'; in (Q) the nonfinite form 'be' is used *as if* it were *finite*. But in (R) the base is chosen as a means of avoiding a declarative sentence in the scope of 'that'; in (R) the base is chosen *because* it is *not finite*. However that may be, I feel safe in assuming logical and philosophical uninterest in this particular employment of 'subjunctive'.

---

[14] Compare Huddleston, *op. cit.*, 80. Also Quirk *et al.*, *op. cit.*, 408.

And the same goes, I take it, for the "formulaic subjunctive", which again uses the base, this time in set formulas which are learned as wholes: '*Come* what may', 'So *be* it', 'God *save* the Queen', '*Suffice* it to say that . . .'.

— And there I rest my case. English boasts nothing that merits the term 'subjunctive' and therefore nothing that merits 'indicative'. English is without distinctions of mood. Those many logicians and philosophers who take the indicative/subjunctive dichotomy as fundamental go wrong at their very first step.

Macquarie University,
Sydney, N.S.W.,
Australia 2109.

© V. H. DUDMAN 1988

MODUS TOLLENS REVISITED

By ERNEST W. ADAMS

I

THE following contextless 'inference' seems to be of *Modus Tollens* form, and since it seems irrational it looks like a counterexample:

If it rained it didn't rain hard. It rained hard. So, it didn't rain.

Of course, if it were a counterexample it would conflict not only with the 'orthodox' material conditional analysis, but with both my probabilistic theory and with Stalnaker's (1968) or Lewis's (1974) 'nearest possible worlds' theories of conditionals. However, closer examination makes it doubtful that it is a counterexample in any straightforward sense.

A possible 'story' of the above reasoning pictures it as a dialogue between two individuals A and B:

A: If it rained it didn't rain hard.
B: It rained hard.
A: So, it didn't rain. (implausible)

Here B appears to implicitly *deny* A's initial statement, and a variant might go:

A: If it rained it didn't rain hard.
B: But it *did* rain hard. (implicitly rejecting A's statement)
A: So, it didn't rain. (implausible)

Assuming that A and B do not accept the same propositions it would be irrational for either of them to accept a conclusion that only follows from both of them together. In fact, given that both A and B would probably accept

If it rained hard it rained.

as a logical truth and that adding that to the other two premisses makes them formally inconsistent, a person could only accept all premisses by accepting a contradiction. But this can't be the whole story.

Many contextless *Modus Tollens* inferences also take dialogue form when they are more fully described. Thus,

> If Jones took chemistry he didn't take physics. He did take physics. So, he didn't take chemistry.

is plausibly represented as a dialogue:

A: If Jones took chemistry he didn't take physics.
B: He *did* take physics.
A: So, he didn't take chemistry.

Even though B's statement no longer formally contradicts A's there is a kind of conflict between them. That is, if A accepted B's statement it would be odd of him to maintain

Jones took physics, but if he took chemistry he didn't take it.

This will be returned to in section IV, but the point to make here is that even if A doesn't maintain the conjunction it might still be rational for him to accept the conclusion. It is easy to see why.

II

Though A doesn't explicitly say so, he is likely to accept the contrapositive of his conditional premiss and reason thus:

A: If Jones took chemistry he didn't take physics.
A: (Also, if he took physics he didn't take chemistry — unspoken)
B: He *did* take physics.
A: So, he didn't take chemistry.

Here B affirms the antecedent of the unspoken contrapositive of A's conditional, which still allows A to arrive at his conclusion, not by *Modus Tollens* reasoning, but by *Modus Ponens*. Note that this is just what would *not* have happened in the first example, because A would have been very unlikely to have accepted the contrapositive in that case. That is, he would have been unlikely to have reasoned:

A: If it rained it didn't rain hard.
A: (Also, if it rained hard it didn't rain — unspoken) etc.

It will be argued in section IV that the 'tacit *Modus Ponens*' analysis extends to other dialogue type inferences that are seemingly of the *Modus Tollens* form, which suggests that when they *are* rational they are really *Modus Ponens* inferences from unstated contrapositive premisses. That reasoners don't always accept the contrapositives of conditionals they affirm explains why this reasoning isn't always 'sound'. This would fit in with both my probabilistic theory and the Stalnaker and Lewis theories, because these theories all reject contraposition as universally valid (in contrast, the material conditional theory does accept its universal validity, which would in turn seem to imply that dialogue type *Modus Tollens* inferences should always be rational).

But if the tacit *Modus Ponens* analysis is right it raises further questions. None of the theories mentioned holds that P is *inconsistent* with 'If C then not P', but not with 'If P then not C', yet there seems to be *some* kind of conflict in one case but not the other, and that needs to be explained. This will be returned to in section IV, but first we will consider an alternative to the tacit *Modus Ponens* analysis.

III

Instead of reasoning by *Modus Ponens* in the second example, A might have argued as follows:

A: If Jones took chemistry he didn't take physics.
B: But he did take physics.
A: If he *had* taken chemistry he wouldn't have taken physics.
A: He didn't take chemistry.

Unlike with the indicative 'If Jones took chemistry he didn't take physics', with which 'He did take physics' seems to conflict, this does not seem to conflict with the counterfactual 'If Jones had taken chemistry he wouldn't have taken physics'. A could coherently affirm

> Jones took physics, but if he had taken chemistry he wouldn't have taken it.

In fact I am inclined to think that most real life reasoning that is represented as being of *Modus Tollens* form has a *counterfactual* conditional premiss rather than an indicative one. Moreover, this would explain why A would draw the conclusion in the second inference but not in the first; namely because he would be willing to affirm the counterfactual in the second case but not in the first. Thus, while A was willing to affirm 'If Jones had taken chemistry he wouldn't have taken physics', he would probably not have affirmed 'If it had rained it *wouldn't* have rained hard' upon learning that it did rain hard. Thus we have an alternative analysis of 'dialogue' *Modus Tollens*: rather than holding that it is really *Modus Ponens* reasoning from unstated contrapositives, we could view it as *Modus Tollens* reasoning from counterfactual conditionals.

I know of no counterexamples to the 'counterfactual' *Modus Tollens* analysis, but it too leads to further problems. Why doesn't the conflict in 'Jones took physics, but if he took chemistry he didn't take it' extend to its counterfactual 'transform', 'Jones took physics, but if had taken chemistry he wouldn't have taken it'? Also, we want to know whether the new analysis is a rival to the tacit *Modus Ponens* theory, or whether both are elements of a larger picture. Probability changes partly account for the reasoning *processes* involved.

IV

Let us make two assumptions: (1) that the probability that C has *after* it is learned that P is the case is equal to the conditional probability of C *given* P, and (2) that this is also equal to the probability of the indicative conditional 'If P then C'. (1) is the thesis that probabilities change by *conditionalization*, which has been a cornerstone of probability theory since the time of Bayes and Laplace. (2) is the *Conditional Probability Thesis*, which is the cornerstone of my probabilistic theory of conditionals. It will be useful to state these formally.

Let probabilities held *prior* to acquiring information be given by the 'prior probability function' PROB_0, and let those held *posterior* to acquiring it be given by the 'posterior probability function' PROB_1. The two theses can then be combined into the double equality:

(1) $\mathrm{PROB}_1\,(C) = \mathrm{PROB}_0\,(C \text{ given } P) = \mathrm{PROB}_0\,(\text{if } P \text{ then } C)$,

Without worrying about the validity of these equalities, let us see how they apply to the phenomena under consideration. We will focus on the propositions $\sim C$ = 'Jones didn't take chemistry' and P = 'Jones took physics', and the probabilities that A attaches to them at different times.

Consider the second example, in which A is represented as coming to believe $\sim C$ = 'Jones didn't take chemistry' after learning P = 'Jones did take physics'. Supposing that to believe something at a time is to attach a high probability to it at that time, for A to come to believe $\sim C$ after learning P is for $\mathrm{PROB}_1\,(\sim C)$ to be high. Substituting '$\sim C$' for 'C' in (1) leads to $\mathrm{PROB}_1\,(\sim C) = \mathrm{PROB}_0\,(\text{if } P \text{ then } \sim C)$, hence if $\mathrm{PROB}_1\,(\sim C)$ is high $\mathrm{PROB}_0\,(\text{if } P \text{ then } \sim C)$ must have been high. Therefore, prior to learning that Jones did take physics A must have believed or 'accepted' the contrapositive conditional 'If Jones took physics he didn't take chemistry', whether he stated it explicitly or not. Hence, if (1) is right the tacit

Modus Ponens analysis must also be right, whatever is the case with the counterfactual *Modus Tollens* analysis.

(1) also explains why 'if C then ~P' = 'If Jones took chemistry then he didn't take physics' might be given up when P was learned. It follows from (1) that when P was learned PROB_1 (if C then ~P) ought to have become equal to PROB_0 (if P and C then ~P). But obviously PROB_0 (if P and C then ~P) ought to have been low, and therefore 'If Jones took chemistry he didn't take physics' ought to become low after it is learned that Jones took physics.

Thus the theses combined in (1) explain the phenomena involved in the tacit *Modus Ponens* analysis of dialogue type *Modus Tollens* reasoning, and so they lend support to the analysis. But that does not necessarily undercut the counterfactual analysis, because the theses apply to indicative conditionals and say nothing about counterfactuals. At one time I had hoped to extend the probabilistic approach to counterfactuals (Adams 1975 Chapter IV, and Adams 1976; see also the *prior propensity theory* of Skyrms, 1980), and I still think this may be possible, but I want to suggest that whether or not the probabilistic approach applies a quite different dimension must also be involved in accounting for counterfactual *Modus Tollens* reasoning.

V

It is plain that explaining the processes that led A to his conclusions is not the same as giving *reasons* for arriving at them. For example, A's coming to believe that Jones didn't take chemistry is explained as being the result of his previously believing the unstated conditional 'If Jones took physics he didn't take chemistry' and his later learning that Jones did take physics. But if A were asked to give his reasons for thinking that Jones didn't take chemistry he wouldn't say 'I thought that if Jones took physics he didn't take chemistry, and then I learned that he took physics'. For A to give reasons for thinking something is not primarily for him to explain the processes that led *him* to think it, but rather to give *someone else* reasons for thinking it. I suggest that that is the function of A's counterfactual assertion 'If Jones had taken chemistry he wouldn't have taken physics', which might also serve to justify A's reasoning. Thus, while the tacit *Modus Ponens* analysis accounts for A's thinking, the counterfactual's role is to influence B's thinking.

If the foregoing is right, or at least on the right track, we are faced with the problem of explaining *how* A's counterfactual utterance together with B's *a posteriori* knowledge (that Jones took physics) could give B reason to think that Jones didn't take chemistry. This is a particular case of what I regard as *the* important problem of counterfactuals, but I have to admit that I have very little to say about its solution. That B must himself be brought to make a *Modus Ponens* inference by being led to think 'If Jones took physics he didn't take chemistry' seems probable, and therefore it is probable that A's counterfactual is intended to give B reason for thinking this. But the 'mechanics' of the belief-inducement are unclear to me, and I can only leave its analysis as an open problem. I will close by explaining why I think the problem is important, and more generally why it is essential to take seriously the 'pragmatic' dimension that comes to the fore when reasoning is represented in dialogue form.

VI

What the dialogue representation stresses that the 'contextless representation' does not is the nature of *the given* in reasoning. Merely writing 'If Jones took chemistry he didn't take physics' and 'Jones did take physics' and asking whether something follows from them gives no clue as to where these items of information came from, who supplies them, and so on. The formalist or abstract semanticist is apt to say that these pragmatic details are irrelevant to the *validity* of the reasoning, but if I am right they *must* be supplied if it is to be understood and evaluated. Only by supplying the details do we see that the reasoning in our examples involves persons with *different* initial beliefs, each of whom supplies a 'given' to the other, and it is a process in which they come to a 'common viewpoint'. But this is itself a logical process, and all reasoning from 'givens' involves something like it. And, if I am right about counterfactuals also, *their* essential logical function, apart from which they cannot be fully understood, is to 'impart' reasons in such processes.[1]

University of California
Berkeley, CA 94720, U.S.A.

© Ernest W. Adams 1988

---

[1] Many of the ideas advanced in this paper were sketched in section IV.1 of Adams (1975). However, there they were illustrated with less perspicuous examples and the points made here were more or less buried in the exposition of a 'hypothesis' concerning counterfactuals which does not figure in the present discussion. I am indebted to Professors William Craig and Brian Skyrms, and to the Editor for very helpful comments and reactions to an earlier draft of this paper. Professor Craig draws attention to the important fact that what I represent as two-person dialogues might equally well be *internal* dialogues, taking place in the minds of thinkers reasoning out the implications of different 'givens'. The Editor points out that the crucial dialogue at the beginning of section II might just as well have gone in three logically quite different directions. Lack of space precludes discussion of these alternatives, but to study them would be to proceed along the same lines as those pursued here.

REFERENCES

E. W. Adams (1975). *The Logic of Conditionals: An Application of Probability to Deductive Logic*, Dordrecht: D. Reidel.
E. W. Adams (1976). 'Prior Probabilities and Counterfactual Conditionals', in W. Harper and C. W. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, Dordrecht: D. Reidel.
D. Lewis (1974). *Counterfactuals*, Cambridge: Harvard University Press.
B. Skyrms (1980). *Causal Necessity*, New Haven: Yale University Press.
R. Stalnaker (1968). 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, Oxford: Blackwell.

VAGUE IDENTITY: EVANS MISUNDERSTOOD

By DAVID LEWIS

GARETH EVANS's article 'Can There be Vague Objects?' ([1]) is over-brief, cryptic, and often misunderstood.[1] As misunderstood, Evans is a pitiful figure: a 'technical philosopher' out of control of his technicalities, taken in by a fallacious proof of an absurd conclusion. Rightly understood, Evans endorses neither the bad proof nor the bad conclusion. Instead he is making a good argument in favour of a very different conclusion. To honour his memory, and to make his point more clearly available, it is worth setting the record straight.

Evans discusses a purported proof that there can be no such thing as a vague identity statement. There are two problems about this proof. One problem is that its conclusion is plainly false. There are vague identity statements. Example: 'Princeton = Princeton Borough'. (It is unsettled whether the name 'Princeton' denotes just the Borough, the Borough plus the surrounding Township, or one of countless somewhat larger regions.) The other problem is that if we understand vagueness as semantic indeterminacy, a deficiency in our language, we can diagnose a fallacy. The proof twice invokes an alleged equivalence between statements of the forms (1) and (2):

(1) it is vague whether . . . *a* . . . ,
symbolized as $V(\ldots a \ldots)$,

(2) *a* is such that it is vague whether . . . it . . . ,
symbolized as $\hat{x}V(\ldots x \ldots)a$.

If vagueness is semantic indeterminacy, then wherever we have vague statements, we have several alternative precisifications of the vague language involved, all with equal claim to being 'intended'. These alternative precisifications play a role analogous to alternative worlds in modal logic. The operator 'it is vague whether...' is analogous to an operator of contingency, and means 'it is true on some but not all of the precisifications that...'. A term like 'Princeton' that denotes different things on different precisifications is, analogically speaking, non-rigid. When *a* is non-rigid, the alleged equivalence between (1) and (2) is fallacious. It is analogous to the fallacious modal equivalence between 'It is contingent whether the number of planets is nine' (true) and 'The number of planets is such that it is contingent whether it is nine' (false), or between 'It is contingent whether the number of planets is the number of planets' (false) and 'The number of planets is such that it is contingent whether it is the number of planets' (true). For a fuller discussion see Thomason [2].

---

[1] The misunderstanding I have in mind can be found in about half of the published discussions of 'Can There Be Vague Objects?' known to me; though never, I think, in the pages of *Analysis*.

The misunderstanding is that Evans overlooks the fallacy, endorses the proof, and embraces the absurd conclusion that there can be no vague identity statements. Besides ascribing folly to a man who was no fool, this interpretation makes nonsense of the title and first paragraph of Evans's article:

> *Can There be Vague Objects?* It is sometimes said that the world might itself *be* vague. Rather than vagueness being a deficiency in our mode of describing the world, it would then be a necessary feature of any true description of it. It is also said that amongst the statements which may not have a determinate truth value as a result of their vagueness are identity statements. Combining these two views we would arrive at the idea that the world might contain certain objects about which it is a *fact* that they have fuzzy boundaries. But is this idea coherent?

How could Evans think that the purported proof — which occupies the rest of the article — addresses his question whether vagueness is due to vague objects, as opposed to vagueness in our mode of describing? A proof that there cannot be vague identity statements would be trouble for the vagueness-in-describing view, no less than for vague objects.

The correct interpretation is that Evans trusts the reader — unwisely! — to join him in taking for granted that there are vague identity statements, that a proof to the contrary cannot be right, and that the vagueness-in-describing view affords a diagnosis of the fallacy. His point is that the vague-objects view cannot accept this diagnosis, because it says that a name like 'Princeton' rigidly denotes a certain vague object. In fact, the vague-objects view does not afford *any* diagnosis of the fallacy, so it is stuck with the unwelcome proof of an absurd conclusion, so it is in bad trouble. (Or better, what is in trouble is the value-objects view combined with the view that vague identity yields identity statements with indeterminate truth value.) On this interpretation, every bit of what Evans says fits into place. However, he has left some important things unsaid.

You might think that charity can be overdone and the textual evidence is inconclusive. One way, Evans comes out saying too much; the other way, too little. What's to choose?

Therefore I end by reporting an exchange of letters in 1978 that ought to settle the matter. A friend sent me a draft taking Evans to task for overlooking the fallacy, endorsing the proof, and embracing the conclusion. I wrote back, hesitantly proposing the interpretation that I have here called correct; and I sent a copy (with my friend's name blanked out) to Evans. Evans replied: 'Exactly! Just so! Yes, Yes, Yes! I am covered with relief that you see so clearly what I was doing . . . and that you were able to ward off the misunderstanding of Anonymous so effectively.'[2]

Princeton University
Princeton, NJ 08544, U.S.A.

© David Lewis 1988

References

[1] Gareth Evans, 'Can There Be Vague Objects?' *Analysis* 38 (1978) 208; reprinted in Gareth Evans, *Collected Papers*, Oxford University Press, 1985.
[2] Richmond Thomason, 'Identity and Vagueness', *Philosophical Studies* 42 (1982) 329–32.

---

[2] I thank Antonia Phillips for her kind permission to quote this passage.

VAGUENESS AND IDENTITY

By B. J. Garrett

THE thesis that there can be vague objects is the thesis that there can be identity statements which are indeterminate in truth-value (i.e. neither true nor false) as a result of vagueness (as opposed e.g. to reference-failure), *the singular terms of which do not have their references fixed by vague descriptive means.*[1] (If this is *not* what is meant by the thesis that there can be vague objects, it is not clear what *is* meant by it.) The possibility of vague objects should not be taken, in itself, to imply the more radical thesis that the identity relation can be one of *degree.* One can hold that the concept of degrees of identity is absurd (how can one thing be more or less identical to another?) *and* that indeterminacy in identity is possible; hence, any incoherence in the idea of degrees of identity does not thereby undermine the idea of indeterminate identity.

---

[1] It seems uncontentious that there can be vague identity statements the vagueness of which is a consequence of the vagueness of their component singular terms — e.g. 'the greatest ruler was the wisest ruler' (Wiggins, p. 174).

I

Gareth Evans attempted to prove, by *reductio*, that vague objects are impossible (ANALYSIS 38.4, October 1978). Evans's argument evidently requires supplementation since, as it stands, it appears to prove the impossibility of *any* indeterminate identity statement and to embody a familiar scope fallacy. A plausible interpretation — or reinterpretation — of Evans's argument, under which the argument is neither fallacious nor has an obviously false conclusion, is the following. Evans is concerned, not with the question of whether there can be vague identity statements, but with the question of whether there can be vague objects, i.e. vague identity statements the singular terms of which do not have their references fixed by vague descriptive means. Evans's proof — on this interpretation — purports to demonstrate that it cannot be indeterminate whether *a* is *b* if neither '*a*' nor '*b*' have their references fixed by vague descriptive means.[2]

Hence, Evans does *not* assume that there are vague identity statements (he simply aims to show, by *reductio*, that there cannot be vague identity statements, the singular terms of which do not have their references fixed by vague descriptive means); his proof is not fallacious; consequently, the issue is *not* how best we ought to respond to a fallacious argument which yields the false conclusion that all vague identity statements are impossible. (The correct diagnosis of the scope fallacy is, of course, available to a believer in vague objects, who does not claim — absurdly — that *all* vague identity statements are reflections of vagueness in reality.[3])

However, on this interpretation, Evans's argument against the possibility of vague objects is, I think, unconvincing. In Evans's proof, (5) does not contradict (1) unless '$\sim(a=b)$' is read as 'it is false that $a=b$'. But the contrapositive of Leibniz's Law entitles us to conclude only that, if *a* and *b* fail to share all their properties, then it is *not true* that $a=b$. To assume that 'it is not true that $a=b$'

---

[2] In what follows '*a*' and '*b*' should be taken as precise designators, the references of which have not been fixed by vague descriptive means.

[3] For this reason, the interpretation offered by Lewis (*Analysis*, above pp. 128-130), despite the endorsement by Evans, just does not present the argument properly. (It may be, of course, that Evans was really congratulating Lewis for realizing that the point of his argument was *not* to show that there could be no vague identity statements.)

imlies 'it is false that $a = b$' — i.e. in effect, to assume, as Evans does, that (5) implies (5') — appears simply to beg the question against a defender of the possibility of vague objects.[4]

II

David Wiggins has recently advanced the following direct argument — not of *reductio* form — against the possibility of vague objects. Wiggins writes:

> Suppose that *a* is *b*. Then, given Leibniz's Law, whatever is true of *a* is true of *b*. Now *a* is definitely *a*. But if so, then *b* is definitely *a*. So by conditional proof and the symmetry of identity, if *a* is *b*, *a* is definitely *b*. ('On Singling Out an Object Determinately', in P. Pettit and J. McDowell (eds.) *Subject, Thought and Context*, Oxford 1986, p. 173)

(We are to take it that '*a* is definitely *b*' implies that '*a* is *b*' is true and hence not indeterminate — otherwise Wiggins's argument would not make contact with its intended target.)

There are, I think, two possible strategies of response to Wiggins's argument. One is to argue that the conclusion (call it (C)) — if *a* is *b*, *a* is definitely *b* — is true, but that its truth is quite consistent with the possibility of vague objects. The second response is to concede that (C) is incompatible with the possibility of vague objects but to argue that, for that very reason, (C) cannot be true.

One source of motivation for the first response is the observation that, in general, the truth of 'if *P* then *Q*' does not imply that '*P*' is either true or false — for example, the conditional 'if this pen is red, it is coloured' may be true even though 'this pen is red' is neither true nor false. Similarly, it may be that (C) is true even though '*a* is *b*' is neither true nor false.

However, there is a disanalogy between (C) and 'if this pen is red, it is coloured'. The latter conditional, we assume, has an indeterminate antecedent and true consequent. But this is not how matters stand with (C). The consequent of (C) cannot have the value *true* if its antecedent is indeterminate in truth-value. If it is indeterminate whether *a* is *b*, the consequent of (C) — which states that *a* is definitely *b* — is just false. But if the antecedent of (C) is indeterminate (hence *not false*) and its consequent *false*, how can (C) be *true*?[5]

---

[4] This response to Evans assumes that, if '*P*' is indeterminate, then '~ *P*' is true (hence, not indeterminate). But this assumption seems correct. If '*P*' is indeterminate, then it is neither true nor false; so it's not true. But that *P* is not true is precisely what '~ *P*' states. Hence, '~ *P*' is true if '*P*' is indeterminate.

A seemingly false consequence of this account is that, if '*X* is bald' is indeterminate, it will be true that *X* is not bald. However, two possible manoeuvres can be made in response. One is to hold that the — tacitly assumed — inference from 'not: *a* is F' to '*a* is not-F' is invalid in the presence of vague predicates. The other manoeuvre is to accept the validity of this inference, but refuse to sanction the inference from '*a* is not-F' to '*a* is definitely not-F'.

John Broome has claimed that a conditional with an indeterminate antecedent and false consequent is *true* (ANALYSIS 44.1, January 1984). If this claim is correct then, despite the disanalogy with the pen-conditional, the first reply to Wiggins goes through. (Of course, if (C) is true, it could feature as the major premise of an argument against the possibility of vague objects. If '*a* is *b*' is indeterminate, it is not true that *a* is definitely *b*; hence, by *modus tollens* with C) as major premiss, it is not true that *a* is *b*, contradicting the assumption that '*a* is *b*' is indeterminate in truth-value. However, if (C) is true, my response to this *modus tollens* argument is the same as that to Evans's argument.)

However, it is plausible to suppose that a conditional with a non-false antecedent and a false consequent cannot be true, but is, rather, indeterminate. (On one such view of the semantics of conditionals, 'if *P* then *Q*' will be true iff either *P* is false or *Q* is true; false iff *P* is true and *Q* is false; otherwise indeterminate.)

If this is correct, what is the best strategy of reply to Wiggins's argument? Clearly, a defender of the possibility of vague objects must deny the truth of (C). (Note: he is not thereby committed to the thesis that (C) admits of degrees of truth — which presumably would mean countenancing degrees of identity. He is committed only to the thesis that (C) is indeterminate in truth-value.) This denial implies, given the validity of Wiggins's argument, that Leibniz's Law, as traditionally formulated, is not true. This may seem to be a *reductio* of the very possibility of indeterminate identity. However, the possibility of vague objects is compatible with the truth of the following conditional (which a defender of the possibility of vague objects may choose to regard as the correct formulation of Leibniz's Law): if it is true that *a* is *b*, whatever is true of *a* is true of *b*, and vice-versa. Arguably, there is no reason to prefer the more traditional formulation of Leibniz's Law *other* than a belief in the impossibility of vague objects, and it is only denial of the modified version — if it is true that *a* is *b*, whatever is true of *a* is true of *b* — which would be absurd.

III

In conclusion: if a conditional with an indeterminate antecedent and a false consequent *can* have the value *true*, then the conclusion of Wiggins's argument is consistent with the possibility of vague objects. If a conditional with an indeterminate antecedent and false consequent *cannot* have the value *true*, then someone who is impressed by apparent examples of indeterminate identity can legitimately subscribe to the following three theses: congruence is a property only of determinate identity; such a restriction leads to no absurdities; and it is a consequence of the most plausible descriptions of certain puzzle cases. Either way, Wiggins's argument is not decisive.

---

[5] I am grateful to David Wiggins for making me aware of this objection in his comments on an earlier draft of this reply.

Lincoln College
Oxford, OX1 3DR

© B. J. GARRETT 1988

A NOTE ON 'MOST'

By JOHN SLANEY

LET 'Most things are F' be written

$$(Mx)Fx$$

and let this be interpreted to be true iff there are more Fs than non-Fs. It is straightforward to extend to a more useful restricted quantifier

$$(Mx.Fx)Gx$$

read 'Most Fs are G' and true iff more things are F-and-G than are F-and-not-G. We may also generalize by allowing arbitrary formulas in place of F and G. The suggested notion is perfectly precise and presents no difficulties at the semantic level.

Some might prefer to think of 'most' as a vague quantifier so that 'Most Fs are G' means something like 'A *substantial* majority of Fs are G'. Personally I consider a simple majority enough for literal truth, but if others feel strongly on the point I shall let it go since for present purposes it is unimportant. Another possible construal of 'most', pointed out to me by Peter Smith, would be to relativize interpretations to systems of 'neighbourhoods' of elements in domains, setting 'Most Fs are G' true iff more Fs are G than are not-G in, say, all sufficiently large bounded neighbourhoods. This would allow us, for instance, to hold with the vulgar that most integers are not prime. I shall not use this note to debate the possible merits of such an alternative account, for my present concern is not to push for a particular regimentation of 'most' but rather to adduce a problem afflicting any reasonable account which generalizes to the infinite case.

One might have hoped that the logic of the quantifier 'M' could be treated syntactically in a fairly standard way, say with finitary introduction and elimination rules in a natural-deduction system with the usual connectives and quantifiers. Such a hope, however, is vain.

Suppose the domain of some discourse is infinite and that the Gs it contains are exactly the Fs together with one individual *a*. Then if most things are G, most things are F, because in the *infinite* case the addition or deletion of just one G cannot make a difference. Analogously, in any infinite domain if most things other than *b* are F then most things are F. For example any set containing most of the nonzero integers contains most of the integers. None of this need be true in finite domains, however: if there are just seven things three of which are F and four, the three Fs and *a*, are G, then most things are G but it is false that most things are F. Now as is well known, we can easily find a set of formulae with no finite model but each of whose *finite* subsets has a finite model. For instance, where n is any finite number, let A_n be a formula asserting the existence of more than n things:

$$\forall x_1 \ldots \forall x_n \exists y(x_1 \neq y \ \& \ldots \& \ x_n \neq y).$$

Then the denumerable set of formulae

$$\{A_1, \ldots A_n, \ldots\}$$

is such a set, and entails the conditional

$$(Mx)(Fx \vee x = a) \rightarrow (Mx)Fx$$

although clearly no finite subset does. It follows that the classical logic of 'most' cannot be captured in any standard system in which consequence is compact (e.g. because all proofs in it are finite).

David Hume Tower,
George Square,
Edinburgh EH8 9JX

© JOHN SLANEY 1988

THE CHAOLOGY OF MIND

By ADAM MORTON

> When an infinitely small variation in the present state may bring about a finite difference in the state of the system in a finite time, the condition of the system is said to be unstable.... the existence of unstable conditions renders impossible the prediction of future events, if our knowledge of the present state is only approximate ... It is a metaphysical doctrine that from the same antecedents follow the same consequents.... But it is not much use in a world like this, in which the same antecedents never again concur. James Clerk Maxwell, 1873

MAXWELL's point is that although identical causes have identical effects (in a deterministic world), it does not follow that even very similar causes have even roughly similar effects. Determinism therefore does not entail predictability. His remarks anticipate a flourishing branch of contemporary physics, which has been called Chaology. Moreover, as Maxwell goes on to point out, his remarks apply in a general way to the special case of the relation between mental and physical states of a person. And when one combines the observation that very similar causes may have very different effects with the realization that similarity is relative to a language one is led to see that though two states of the nervous system may be very similar when described neurologically, the differences in their effects may make them very different when described as states of mind. This is in effect a formulation of minimal materialism, or the token identity theory of mind-brain identity.

My aim is to show that the resemblance between unpredictability in deterministic physical systems and token identity theories in the philosophy of mind is no accident. I first show that chaology provides a model for the relation between two vocabularies postulated in token identity theories, and then I argue for the more speculative claim that deterministic unpredictability may be the reason why token identity theories are true.

I Models of Token Identity

Theories such as those of Lewis and Davidson, which postulate that tokens of mental and physical types are tokens of physical, particularly neural, types, can be characterized in terms of three central claims and two less central claims about the relations between two vocabularies, ϕ and ψ, meant to be our stock of physical and mental words respectively. I shall take the tokens which are described by these vocabularies to be states of mind and brain. Similar claims could be made for physical and mental events and properties.

(i) Token identity: Every ψ state is a ϕ state.
(ii) Type non-identity: Classes of states as picked out by the ϕ vocabulary need not coincide with classes picked out by the ψ vocabulary, and vice versa.
(iii) Maxwell's point: Arbitrarily similar ϕ (ψ) states can be very different ψ (ϕ) states.

To these I shall add two more claims. The first is meant to capture part of what Davidson calls the anomalous nature of the mental. And the second gives a constraint on too extravagant formulations of anomalousness.

(iv) Anomalousness: There are no statable laws which reduce all ψ descriptions to ϕ descriptions.
(v) Qualified correlation: In a number of limiting cases one can deduce the ϕ description from the ψ description.

I shall say that vocabularies satisfying these conditions are 'S-related'. (S for supervenience.) The attractions of talking this way in the philosophy of mind are familiar. It would be rather disturbing, though, if mental and physical vocabularies were the only example of vocabularies related in this way.

Analogies of the first two conditions above are satisfied by the standard example, that of symbol-tokens classified in various ways. But even for these two conditions the analogy to mental and neural processes is not very good, for the objects of the descriptions are objects rather than states or events. And the example does not satisfy the remaining three conditions in any very clear way.

Here is a better example. The two vocabularies describe the states of a physical system and the laws which govern its evolution. The "ϕ" vocabulary describes the system at particular times and the laws of its evolution, attributing to it particular precise states ('positions in phase space') and simple classes of states ('geometrically simple regions of phase space'), and also formulating the laws by which one precise state evolves to another. I shall call this the vocabulary of momentary states. The "ψ" vocabulary describes possible patterns of evolution of the system through time, attributing to it geometrically simple trajectories through actual or phase space. A description of the system as following a straight or a cyclic path belongs to the ψ vocabulary. So does a description of it as tracing a path depicting the Eiffel tower or as having a prime number of left-hand turns. I shall call this the vocabulary of trajectories.

It is important to see that the contrast between these two vocabularies is not simply that between states-at-a-time and laws, on the one hand, and trajectories on the other hand. For the distinction is one of descriptive vocabularies rather than of states described, and some states-at-a-time may be defined in terms of trajectories. For example a description of the form 'will begin in *n* seconds time to follow a trajectory of kind *T*' or 'might in *n* seconds follow a trajectory of kind *T* if...', describes a state-at-a-time, but in ψ vocabulary terms. And, similarly, a trajectory-based description of the form 'the path followed according to laws *L* from initial state *S*' gives in the ϕ-vocabulary a description of a trajectory.

In the cases I am concerned with the system is deterministic but chaotic. That is, the laws governing the evolution of the system are stated by a theory *T* and a function *D* with two features. (a) Determinism: *T* entails that $D(S, t, t')$ is the exact state of the system at any time t, given as a function of its state S at any earlier time t'. (b) Chaos: with increasing t D takes simple sets of precise states ('geometrically simple regions of phase space') to complex sets which cannot be described by the ϕ vocabulary ('geometrically complex regions of phase space: in the limit regions with fractal boundaries'). In effect, as time goes on sets of arbitrarily similar initial states evolve to more and more diffuse sets of very dissimilar states.

There are many such systems. A simple pendulum swinging from a very slightly oscillated support shows chaotic motion: an infinitesimal alteration of its initial conditions can cause it to follow an entirely different trajectory. The flow of a fluid in a container or around an obstacle is chaotic as soon as any degree of turbulence enters the picture. The motion of colliding billiard balls, those models of determinism, is chaotic given even the slightest gravitational influence from beyond the table.

Given such a system, the descriptions of its states in terms of the vocabulary of momentary states and that of trajectories are S-related. The five conditions are satisfied, as follows:

(i) Token identity: The system at any given time is in a particular token state (a particular point in phase space), which though sometimes definable in terms of trajectories always has a definition as a momentary state.
(ii) Type non-identity: Many geometrically simple classes of trajectories define geometrically complex classes of initial states, and vice versa.
(iii) Maxwell: States which are very similar when described in momentary state terms can be very different when described in terms of the trajectories they lead to.
(iv) Anomalousness: The laws that govern the evolution of the system do not entail laws relating classes of trajectories to classes of initial states.
(v) Qualified correlation: Some classes of trajectories are definable in terms of initial conditions, so some descriptions in terms of trajectories can be unpacked into momentary descriptions.

Exactly similar remarks could be made about trajectories defined in terms of the two vocabularies.

II DISPOSITIONS, TRACKING STATES

The relation between initial conditions and trajectories can thus model the relation between physical and mental states postulated by token identity theories. The model is more interesting than that of types and tokens of written characters, for it concerns physical states and their causal properties. As a result it may be of use in exploring the ways in which mental life and neurological functioning may be related. In particular, it seems a good model to use when trying to make precise the ideas of token states and events – essential to most sophisticated materialisms.

But, still, the model may be just a model, whose details reveal nothing about what is modelled. I cannot show conclusively that it is not. But this section and the next are meant to describe a more interesting possibility, that the model may have its roots in some deep and characteristic facts about mind.

To begin, is there any reason to believe that those states of the nervous system which we describe in mental terms have any relation to trajectory-based descriptions of physical systems? I think the case can be made for two kinds of psychological state: dispositions and tracking states.

I shall take dispositions to be states of mind definable in terms of what an agent would do and experience in various possible circumstances. Dispositional descriptions of states of mind are defined in terms of possible patterns of development of actions and experiences-at-a-time, and so they are clearly "trajectories" in relation to instantaneous states of mind. By 'tracking states' I mean those states of mind which represent particular physical objects and facts in such a way that the state may be said to 'track' the object or fact: if the object or fact were to change in specific respects, so would the state. Various tracking states have been studied by Grice, Unger, Goldman, Morton, and Peacocke. Tracking states are also in an obvious way 'horizontal' to momentary states of mind. That is, they are a kind of counterfactual disposition: if the environment [or intention] were different in respect *X* then the state of mind [or action] would be different in respect *Y*. They too are thus trajectory-based. But there is an important difference: the relevant causal influences derive as much from the environment as from the mind or nervous system itself. These are only two of the many kinds of mental state that are plausibly seen as trajectory-like. For states of mind usually concern not what is going on in the person at an instant but what pattern of action, feeling, or thought is being played out over a span of time.

At this point I must assume a limited token identity of mind and brain, in effect assumptions (i) and (iv) above. Given these, it follows that each state of mind at a time can be identified with a state of the nervous system. *Then* dispositions, related to them as trajectories to momentary states, are in fact trajectory-based states of a physical system, the brain. And tracking states are trajectory-based on momentary patterns of firing of neurons we see that the states of the nervous system studied by neurophysiology are based on momentary patterns of firing of neuros we see that neurological and psychological descriptions are ϕ and ψ vocabulary descriptions of a physical system as described in the previous section. And therefore *if* the system is deterministic but chaotic, the token identity is accompanied by a type non-identity.

Is the system chaotic? Most sufficiently complex physical systems are, and the complexity of the nervous system is astounding. And when we consider systems consisting of the nervous system and parts of the environment, chaotic evolution becomes inevitable. The appearance of chaotic evolution in the working of the nervous system is so safe an assumption, in fact, that what needs explanation is its stability, the extent to which later states of mind and brain *are* predictable in terms of earlier ones. People remember much of what they experience, carry out chains of reasoning in a determinate way, and so on. Not without exception, but regularly enough for it to be remarkable that a system of so many parts interacting in so many ways and subject to so many forces should be even to this degree predictable.

Now the really interesting questions open up, and the answers to them are not at all obvious. What kinds of chaos do neural systems exhibit? Which of these kinds are relevant to states of mind? Answering these would take us a small way towards answering the most basic and important question about mind and body: which neural kinds are mental kinds?

III CHAOTIC EXPLANATION

There are many ways in which a physical system can be unpredictable. One easily imagined way is illustrated by a sphere falling onto a sharp spike and being deflected to right or left until it meets another spike further down, and so on. Here although the system is deterministic the dependence on the exact point of impact between sphere and spike is so delicate that the system evolves as if it made a random choice at each impact. This is one kind of deterministic unpredictability. Many other physical systems although physically very different have just this kind of unpredictability. But it is fairly clear that it is not this kind of unpredictability that we can expect to find in the psychologically relevant functioning of the nervous system. It doesn't capture the mix of stability and chaos typical of mind, either in its internal evolution or in its response to the enviroment.

To make things a bit more definite, suppose that one was trying to give necessary conditions for a state of the nervous system to be a tracking state of mind. The first thing to note is that tracking states exhibit to a high degree the paradoxical predictability-within-chaos of mind-bearing brains: as the state of some external object changes (for example) the perceptual state of the observer changes accordingly, but between the external object and the perceptual state there is no direct linkage but a complicated, unpredictable and easily diverted network of causal paths. This seems to me to fit very naturally ito what is becoming the standard typology of chaotic systems. They are usually classified in terms of the connections between their 'control parameters' and their 'attractors'. Control parameters describe causal influences from outside the system and quantifiable differences from physically similar systems exhibiting different patterns of evolution. Attractors are regions of phase space into which all sufficiently near states of the system eventually evolve. The system is classified in terms of the way in which the values of the control parameters affect the resulting set of attractors.

Tracking states fit easily into this style of analysis. The control parameters are set (in part) by the states of the environment, e.g. by the location of a perceived object. The attractors of the system are stable perceptual states, for example those registering the locations of perceived objects. Thus each perceptual state corresponds to a large class of neural states — the attractor — whose boundaries are often very complex. Given a specific value of the control parameters and a nervous system in a wide but not universal set of initial states — e.g. an object perceived under normal conditions by an observer — the system will settle down to a state within the corresponding attractor — e.g. the appropriate perceptual state will result.

Putting things this way does not give any laws connecting perceptual situation and perceptual state, or help determine which neural states are perceptual tracking states. What it does is to reveal that the explanatory strategy employed in chaology subsumes as a special case differential explanation, the pattern of explanation appropriate to tracking states. And this does suggest — only suggest — that the vocabulary and methods being developed to classify and understand chaotic systems may be part of what we need to understand how states of the nervous system can be states of mind. For one thing, in producing explanations and something like predictions within a context of essential unpredictability chaology wrestles with some of the problems facing psychological explanation. And faced with the general problem of relating patterns of activity over time — as many psychological states surely are — to instantaneous states, in the presence of a combination of rough overall predictability and small-scale chaos, the study of the behaviour of essentially unpredictable physical systems is too promising a tool to ignore.[1]

---

[1] The Maxwell quotation is from a lecture reprinted in Campbell and Garnett [1]. It was brought to my attention by Professor Michael Berry, the coiner of the term chaology, whose writings and conversation introduced me to the subject. Tritton [8] and Crutchfield and others [3] explain the main points of chaology. For a thorough exposition see Schuster [7], and for a suggestion of its importance to philosophy see Hunt [6]. The definition of determinism should really be vastly more involved: see Earman [4]. The possible use of my model to clarify the ideas of a state and an event was suggested to me by Chris Allman's PhD thesis (Bristol 1987). Ideas like those of my last section are found in Harth [5], pp. 208–14. For more on neural states as points in a phase space see Churchland [2]. I owe the term 'minimal materialism' to Pascal Engel. David Hirschmann has given me much advice and encouragement, Marie McGinn caused some last-minute revisions, and the Editor pointed out a serious mistake.

9 Woodland Road,
Bristol, BS2 8XY

© Adam Morton 1988

REFERENCES

[1] Lewis Campbell and William Garnett, *The Life of James Clerk Maxwell*, London: Macmillan, 1882.
[2] Paul Churchland, 'Phase-Space Representation and Coordinate Transformation', *Behavioral and Brain Sciences* 9 (1986) 93–4.
[3] James P. Crutchfield, J. Doyne Farmer, Norman H. Packard, and Robert S. Shaw, 'Chaos', *Scientific American* 255 (1986) 46–57.
[4] John Earman *A Primer on Determinism*, Dordrecht: Reidel, 1987.
[5] Erich Harth *Windows on the Mind*, Brighton: Harvester, 1982.
[6] G. M. K. Hunt 'Determinism, Probability and Chaos', *Analysis* 47.3 (June 1987) 129–33.
[7] Heinz Georg Schuster, *Deterministic Chaos*, Weinheim: Physik-Verlag, 1984.
[8] David Tritton 'Chaos in the swing of a pendulum', *New Scientist* 111 (1986) 37–40.

THE WATERFALL ILLUSION

By TIM CRANE

IF you stare for a period of time at a scene which contains movement in one direction, and then turn your attention to an object in a scene which contains no movement, this object will appear to move in the opposite direction to that of the original movement. The effect can be easily achieved by attaching a piece of paper with a spiral drawn on it to the spinning turntable of a record player, and then turning the turntable off while continuing to look at the spiral (see Frisby 1979, pp. 100–101 for a detailed description of how to bring this about). But the illusion of movement can also occur when looking at a waterfall, for instance, and turning one's attention away from the waterfall to a stationary object such as a stone; hence its name — the 'Waterfall Illusion'.

The effect is quite striking, and not difficult to achieve. But the above description is not quite right. For although the stationary object *does* appear to move, it does not appear to move relative to the background of the scene. That is, there is a clear sense in which it also *appears to stay still* (see Blakemore 1973, p. 36). There is a distinct appearance of lack of motion as well as motion. Understandably enough, many find this aspect of the illusion quite extraordinary; John Frisby writes that

> although the after-effect gives a very clear illusion of movement, the apparently moving features nevertheless seem to stay still! That is, we are still aware of features remaining in their 'proper' locations even though they are seen as moving. What we see is logically impossible! (Frisby 1979, p. 101)

Presumably what Frisby thinks is logically impossible is that something could be both moving and not moving at the same time; and this claim must be correct. I will argue that this aspect of the Waterfall Illusion poses a problem for a familiar claim about perception: the claim that the content of perceptual experience is composed of *concepts*.

Of course, what one makes of such cases as the Waterfall Illusion partly depends on one's methodology. One could dismiss such cases as uninteresting exceptions, to be dealt with once the truth about normal perceptual functioning has been found. Perhaps this is the right way to treat illusions. But my own sympathies lie with Helmholtz:

> The study of what are called illusions of the sense is, however, a very prominent part of the psychology of the senses; for it is just those cases which are not in accordance with reality which are particularly instructive for discovering the laws of those processes by which normal perception originates. (Quoted in Coren and Girgus 1978, p. 9)

On this view of the significance of illusions, the study of illusions could be compared to pathology: one studies the body's abnormal functioning in order to understand normal functioning. Or — closer to home — it could be compared to Russell's idea that one uses logical puzzles (about, e.g., non-referring singular terms) to test the adequacy of a logical theory (see Russell 1957, p. 47). I shall adopt Helmholtz's advice, and see what we can learn about perception from the Waterfall Illusion.

The problem is: what is the content of such a visual experience? Note that the problem concerns the *content* of the experience, not the *attitude*. We are familiar (from discussions of phenomena such as the Muller-Lyer Illusion) with the fact that illusions can occur even when the victim of the illusion knows the truth about what he is seeing. And there is an extensive discussion of whether this fact can be accounted for by a theory which claims that perception is a sort of belief. (Armstrong 1969 argues that it can, Jackson 1977 and Fodor 1984 argue that it cannot.) I cannot enter this debate here, and I do not need to; for *whatever* the fate of the idea that perceptions are beliefs, there is still a problem about characterizing the *content* of an experience like the Waterfall Illusion. In fact, it seems to me to be implausible that the victim of the illusion, knowledgeable or not, would judge the stationary object to be moving — though he might judge that he is drunk (a similar effect can be had by playing the child's game of spinning oneself around and then trying to stay still). But this is not directly relevant to the present point; we are concerned in this paper with contents, not attitudes.

The crucial difference between the Muller-Lyer Illusion and the Waterfall Illusion is this. The Muller-Lyer Illusion presents a conflict between *two* intentional states: the state of believing that the lines are the same length, and the state of the lines *looking to be* different lengths (I leave open the question as to whether this is a 'suppressed inclination to believe' as the belief theory of perception holds). The Waterfall Illusion, however, presents a contradiction in the *one* content of one attitude. The viewed object seems to be both moving and not moving at the same time. This is the natural way to describe how things seem; unlike the Muller-Lyer illusion, the content of the *experience itself* is contradictory.

Now if the content of this experience is indeed contradictory, then a significant problem arises for those who claim that the contents of experiences are individuated according to certain familiar principles. We may call these principles 'Fregean' because of their origin in Frege's theory of Sense and Reference. Frege introduced a criterion of difference for senses which distinguished senses as the *cognitive value* of expressions; for example, the criterion of difference for the senses of sentences may be expressed as follows:

(I) for any thinker *A*, and any sentences *S* and *S'*, if *A* understands *S* and *S'* and accepts *S* as true while not accepting *S'*, then *S* and *S'* have different senses.

(I) allows subjects to take different attitudes to sentences with the same truth-conditions; the notion of sense is introduced to explain this possibility. So (I) is meant to individuate the senses of sentences (their *contents*) finely enough to rule out the unnecessary ascription of contradictory beliefs to rational subjects. *De re* considerations aside, we do not want to credit Ralph with the belief that Ortcutt is and is not a spy. So (I) is invoked to explain why it is that Ralph does not have contradictory beliefs in this situation, and therefore how identity statements can be informative. The principle is thus meant to provide a test for the cognitive significance of senses; it does not tell us what this is, but it tells us when it is distinct.

Clearly, similar principles apply to the constituents of contents: the senses of singular terms and predicates. If we call the sense of a predicate a *concept* (departing, of course, from Frege's own use of that term) we can introduce a criterion of difference for concepts inspired by (I):

(II) *F* and *G* are different concepts if it is possible for a subject to rationally judge, of an object *a*, that *a is F* and that *a is not-G*.

Like (I), this principle allows the possibility that a rational subject can apply incompatible concepts to the same object, by taking these concepts to differ in cognitive significance.

One writer who has urged that we individuate contents and their constituents in this way is Christopher Peacocke (see, e.g., Peacocke 1984, p. 365). Now Peacocke has also suggested that we use these principles in individuating the contents of perceptual experience (see his 1983; but for an important reconsideration not related to the problem discussed in this paper see his 1986, pp. 9–11). In *Sense and Content* he claims that

> it is a conceptual truth that no-one can have an experience with a given representational content unless he possesses the concepts from which that content is built up. (1983, p. 19)

If the concepts which build up the representational content of an experience are individuated along the lines of principle (II), then we arrive at a further principle:

(III) *F* and *G* are different perceptual concepts if it is possible for a subject to have (at the same time) an experience with the content that *a is F* and an experience with the content *a is not-G*.

It follows from this that one cannot have an experience with contradictory representational content. This thought seems to have struck Edward Craig, who writes (in defence of his claim that experience 'belongs to the category of judgement') that one cannot see a Necker Cube looking one way, and then imagine it looking *at the same time* the other way (see Craig 1976, p. 13).

But, as we have seen, the Waterfall Illusion is precisely a case where a scene looks as if it is both one way and another (incompatible) way at the same time. Craig's point is no doubt true of the Necker Cube (and indeed, of most normal perception). But the Waterfall Illusion is a vivid counterexample to a generalization of the point. For what it suggests is that a subject can be ascribed an experience with a contradictory content. But if principle (III) is central to the individuation of perceptual concepts (as it seems to be) then the Waterfall Illusion is, surprisingly, a counterexample to the thesis that concepts are involved in the content of perceptual experience. For the subject can see that the stone (for example) is moving and that it is not moving. But this contravenes (III), since we know that there is only one concept being exercised: 'moving'.

It seems to me, then, that this illusion presents a problem for those views (such as Peacocke's (1983) or Craig's (1976)) which treat the content of perception as conceptual. The argument is not decisive — there are various ways in which one could resist it. One response would be to deny that the two contradictory halves of the content are *really* simultaneous (so that the situation is rather like a very rapid switching between the two 'aspects' of the Necker Cube). But this response seems, in the face of the phenomenology, unduly stipulative.

Another response would be to insist that the two halves of the content are not really contradictory, because the concepts involved in each half are distinct. But how plausible is this response? If the concepts in the two halves of the content are not the same, it is hard to see precisely why the illusion presents this apparent contradiction. Indeed, it is difficult to say exactly what the two distinct concepts *are*; perhaps in the case of looking at a waterfall and then a stone, they might be 'intrinsically moving' (whatever this may mean) and 'not moving relative to other objects'. But in the case of looking at the stationary turntable, the concept 'not moving relative to other objects' cannot be the appropriate one, for the position of the spiral on the turntable relative to other objects is *irrelevant* to the production of the illusion. (The turntable itself could be moving through space, and the illusion would still occur.) The problem for the defender of concepts in perception is to say exactly which concepts are being exercised here, without denying that there is a conflict in how the perceived scene seems to be.

Finally, one might dispense with principles (II) and (III) and try to find some other constraint on the identity of concepts. This would surely be an excessively severe reaction. I suggest that we should leave the Fregean constraints where they belong — with the 'higher' cognitive faculties of judgement, belief and thought, those faculties which are governed by the 'constitutive ideal of rationality' (Davidson 1982, p. 223). Perhaps the fact that perceptual contradictions can occur (however rarely) in the minds of otherwise rational subjects suggests that perceptions are not subject to all of the principles which govern the operations of the higher faculties; perception is, perhaps, a 'sub-rational' process. And this may seem to be supported by Fodor's view that the operations of the perceptual system are 'informationally encapsulated'; that is, the informational content of perceptual states cannot be affected by the contents of states in 'central mind' (Fodor 1984, pp. 64–86). Of course, Fodor's theory of the 'modularity' of visual perception does not on its own explain how there can be perceptual contradictions. But it does suggest an explanation of why perception may not be a wholly rational process, and thus why it is peculiarly vulnerable to phenomena like the Waterfall Illusion.[1]

Peterhouse,
Cambridge, CB2 1RD

© TIM CRANE 1988

REFERENCES

D. Armstrong 1969, *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul.

C. Blakemore 1973, 'The Baffled Brain', in E. Gombrich and R. Gregory (eds.) *Illusion in Nature and Art*, London: Duckworth.

S. Coren and J. Girgus 1978, *Seeing is Deceiving*, London: Wiley.

---

[1] Many thanks to Jeremy Butterfield, Hugh Mellor and the Editor for useful comments.

E. J. Craig 1976, 'Sensory Experience and the Foundations of Knowledge', *Synthese* 33, pp. 1–24.

D. Davidson 1982, *Essays on Actions and Events*, Oxford: Oxford University Press.

J. Fodor 1984, *The Modularity of Mind*, Cambridge, Mass.: MIT Press.

J. P. Frisby 1979, *Seeing*, Oxford: Oxford University Press.

F. Jackson 1977, *Perception*, Cambridge: Cambridge University Press.

C. Peacocke 1983, *Sense and Content*, Oxford: Oxford University Press.

C. Peacocke 1984, 'Colour Concepts and Colour Experience', *Synthese* 58, pp. 365–81.

C. Peacocke 1986, 'Analogue Content', *Proceedings of the Aristotelian Society Supplementary Volume* 60, pp. 1–17.

B. Russell 1957, 'On Denoting', in R. C. Marsh (ed.) *Logic and Knowledge*, London: Allen and Unwin.

CRANE'S WATERFALL ILLUSION

By D. H. MELLOR

T. M. CRANE ('The Waterfall Illusion', ANALYSIS above, pp. 142–7) thinks his eponymous illusion teaches us something about perception. I agree, but not with what it's taught him, namely that perceptual experiences aren't 'composed of concepts' in the sense that beliefs are. Take the concept of motion (*F*) and the belief that something moves. No one who lacks that concept can have that belief, because the belief in some sense contains the concept. What the sense is, and whether concepts are conceptually prior to beliefs, are good questions, but ones I needn't go into. The question here is whether perceptual experiences also contain concepts in this sense, whatever it is. Crane thinks the Waterfall Illusion (WI) experience shows that they don't. I think it shows no such thing.

In the Waterfall Illusion an object *a* apparently looks both to move and not to move. I say we need the concept *F* to have that experience, just as we need it to believe that *a* is *F*. For how, unless it contained that concept, could 'the content of the *experience itself* [be] contradictory', as Crane says it is (p. 144)? What is the contradiction, is not *Fa* & ~*Fa*? Crane doesn't say. Nor does he say how, without containing the concept *F*, the WI experience can include believing that *a* looks to be both *F* and not-*F*, which it evidently does.

Why then does Crane deny that in the WI experience *a* is conceived both to be and not to be *F*? His denial is based on his criterion (II) of difference for concepts, inspired by his Fregean criterion of difference (I) for the senses of sentences. And from (II) he derives the following principle:

> (III) F and G are different *perceptual** concepts if it is possible for a subject to have (at the same time) an experience with the content that *a is F* and an experience with the content that *a is not-G*. (p. 145, *my italics.)

If the content of the WI experience is $FA \mathbin{\&} \sim Fa$, (III) is false. Whereas (III)'s analogue for belief is very plausible, at least for conscious belief:

> (IIIB) F and G are different *belief* concepts if it is possible for a subject to have (at the same time) a belief with the content that *a is F* and a belief with the content that *a is not-G*.

As it stands, of course, (IIIB) purports only to be a criterion of difference for belief concepts, as (III) does for perceptual concepts. But the concept F is the same whether it is a constituent of a belief, a fear, a desire — or a perceptual experience. The italicized words '*perceptual*' and '*belief*' should therefore be deleted from (III) and (IIIB). In fact for present purposes (IIIB) may be simplified further to

> (B) No one can have a conscious belief with the content $Fa \mathbin{\&} - Fa$,

and (III) to

> (P) No one can have a perceptual experience with the content $FA \mathbin{\&} \sim Fa$.

What the WI experience seems to show is that even if (B) is true, (P) isn't. But Crane thinks (P) must be true of any experience that contains the concept F, so the WI experience can't really have the content it seems to have. But why not? After all, many of (B)'s other analogues seem equally false: we can, it seems, easily fear or desire (for one reason) that a is F while fearing or desiring (for another) that it isn't. That doesn't make anyone deny that those fears and desires really contain the concept F. Why then deny that the WI experience does so too?

A reason emerges when we ask why contradictory beliefs are more problematic than contradictory fears and desires. The answer is, of course, that belief, unlike fear and desire, aims solely at truth, in the sense shown by Moore's paradox (here contraposed): it's only absurd to say something of the form 'I X that p but it isn't true' for all p when 'X' is or entails 'believe'. To believe a proposition p is always just to think it true, which isn't generally true of other propositional attitudes. No one fears or desires what they think they already have — they may fear or desire its continuance, but that's another matter. So 'I fear/desire that p and it isn't true' is hardly ever an absurd thing to say. That's why contradiction in the content of a fear or desire is usually no more problematic than it is in the content of a disbelief.

But perceptual experience isn't like fear or desire. Like belief, although it may be false, it seems to aim solely at truth: its function, after all, is to tell us how the world truly is. So why isn't (P) as true as (B)? Perhaps it is — because (B) is false too, since we *can* have contradictory beliefs, Moore's paradox notwithstanding. Suppose I believe that all even numbers are the sum of two primes, and you don't. One of us has a contradictory belief: but will anyone therefore deny that it contains the concepts of even and prime numbers? Well, maybe some philosophers would deny that but we can leave that question open. For whether or not we can consciously believe so complex a contradiction, we can hardly consciously believe one as simple as *Fa* & ~ *Fa*, and certainly not while being conscious that it *is* a contradiction: because to be conscious of that is to see that it can't be true. Read that way, (B) is certainly true.

Yet the WI experience seems to show that, read the same way, (P) is false. In it we do seem to be conscious of having an experience with a contradictory content: *Fa* & ~ *Fa*. But how can a perceptual experience have such a content if, like belief, it aims solely at truth? Crane thinks it can't. I think it has to: so I must say how.

First, as Crane has pointed out to me, the sense in which perceptual experience aims at truth isn't captured by the analogue of Moore's paradox. It isn't absurd to say 'I see that *a* is *F* but it isn't' (ignoring the trivial 'achievement' sense of 'see' and the sense in which it entails 'believe'). But this just follows from the fact that we don't always believe what we see: a fact shown as much by the Muller-Lyer as by the Waterfall Illusion. But then in what sense does perceptual experience aim at truth?

The obvious answer is one that Crane mentions but doesn't endorse: perceptual experience is or entails an 'inclination to believe' (p 144). The experience has the content of the belief, but may not produce it, because it may conflict — via (B) — with an existing belief. The inclination to believe is still there, but is suppressed, as in the Muller-Lyer Illusion. But that, as Crane says, can't be how the Waterfall Illusion works, because there we are conscious of a contradiction in the supposed content of the experience itself. That experience can't be a suppressed inclination to believe that content, since we can't be inclined to believe so patent a falsehood in the first place.

We could, however, be inclined to believe that *Fa* while also being inclined to believe that ~*Fa*. And that, I submit, is what happens in the Waterfall Illusion. There isn't simply, as Crane claims, 'a contradiction in the *one* content of one attitude' (p. 144). Rather we are conscious of seeing that *a* moves while also seeing that it doesn't. One of these two perceptual experiences gives us the corresponding belief, say that *a* doesn't move, which then suppresses the rival inclination to believe that it does.

The real question is: what produces these conflicting inclinations, to believe that *a* moves, and that it doesn't, at the same time? I think the answer is that motion includes two distinct concepts, each with its own perceptual mechanism: mov*ing* (motion at an instant), and *having* moved (being in different places at different times). We know we can see that *a* has moved without ever seeing it moving; and the converse is also true. We know we can see something moving without seeing that it's moved, since, as we know, a moving object can catch our eye when a stationary one wouldn't. In other words, we have an independent perceptual mechanism for seeing things moving. It may, of course, take time: maybe we can't see absolutely instantaneous motion. But the mechanism can still show us perceptibly instantaneous motion: that is, motion over an imperceptibly small time span, which we conceive to be instantaneous. And when this mechanism (rightly or wrongly) shows us *a* moving, it will naturally incline us also to believe that *a* has moved: since moving for any finite time entails having moved. So it also inclines us to believe simply that *a moves*: it gives us a perceptual experience whose content is *Fa*. And normally, of course, when that experience is veridical, it will be confirmed by the mechanism which shows us that, over a perceptible time span, *a has* moved. But it needn't be so confirmed; and in the Waterfall Illusion, it isn't: the two mechanisms simultaneously give us conflicting perceptual experiences.

In short, not only must we credit the Waterfall Illusion experience with containing the concept *F*, in order to explain its contradictory content, we can. So we should. It is an illusion to suppose, as Crane does, that this illusion shows perceptual experience not to be composed of concepts.[1]

Darwin College,
Cambridge CB3 9EU

© D. H. MELLOR 1988

---

[1] Notwithstanding our disagreement, the above owes much to Mr Crane's replies to my earlier criticisms of his paper.

CONCEPTS IN PERCEPTION

By TIM CRANE

I CAN agree with much of what D. H. Mellor says in his response to my paper ('Crane's Waterfall Illusion' ANALYSIS, above pp. 147–152). I can agree that perception in some sense 'aims' at truth, that its function 'is to tell us how the world truly is' (p. 149). I can agree that perception *normally* inclines us to believe that what is perceived is the case — we normally believe what we see. And I can agree with Mellor's proposal that the ability to perceive motion might be the product of two distinct but related mechanisms. The disagreement between us is over the tentative conclusion I drew from my reflections on the Waterfall Illusion: that the content of perception is not composed of concepts. Mellor seems to be sceptical about the thesis that there could be a state with a content which is not composed of concepts. On the face of it, there is something odd about this thesis, since concepts are often *defined* as the constituents of contents. If I am to defend my reading of the Waterfall Illusion, I must try and make this thesis more palatable.

The substance of Mellor's criticism is that the content of the illusion should not be constructed (as I construe it) as *Fa* & ~ *Fa*, but as the conjunction of two distinct contents, *Fa* and ~ *Fa*, each of which the perceiver is inclined to believe. One of these inclinations, as it were, 'wins out', presumably because of its support from other beliefs and perceptions. In my paper I argued that the problem for those who believe that the content of perception is conceptual arises because the Waterfall Illusion is an experience with a single contradictory content. On Mellor's interpretation of the illusion, this problem simply does not arise. His concern is to preserve the equation of perception with an inclination to believe, in order to capture the sense in which perception 'aims' at truth, for this sense (while distinct) cannot be unrelated to the sense in which belief aims at truth.

But I doubt whether this is the whole story. How can the notion of an inclination to believe capture what is distinctive about the content of perception? In particular, how can it explain the general fact (of which the Waterfall Illusion is an illustration) that, unlike beliefs, the contents of one's perceptual states cannot be *revised* in response to further evidence? If perceptions were simply inclinations to believe, one would expect that the inclination which 'loses' (in our case, that *a* is moving) would no longer be present to the mind. This is surely so with *non-perceptual* inclinations to believe: once conclusive evidence is presented against one's (non-perceptual) inclination to believe that *p*, it would be irrational to continue to be inclined to believe that *p*. But in the case of the Waterfall Illusion, the (alleged) inclination to believe that *a* is moving remains even after conclusive evidence has been assembled against it. So the question remains: what distinguishes the perceptual inclination from the non-perceptual one?

I can concede that perceptions *normally* involve an inclination to believe. But to say this is just to say that perceptions are generally reliable in the production of beliefs; and this is not news. The real question, as Mellor says, is how phenomena like the Waterfall Illusion come about, and what they tell us about the distinction between perception and belief. Mellor has a proposal about this, which (in part) I find plausible: the perception of motion is the result of two mechanisms, corresponding to the ideas of *moving* and *having moved.* Mellor thinks that these two mechanisms deliver perceptual experiences whose contents are composed of concepts; I do not. But if I am to be right about the content of perception, then I have to explain *why* not; that is, I have to explain how a perception that *p* can have the content *p* without being composed of the concepts that the belief that *p* is.

This explanation can only be complete when an adequate account of the possession of concepts is given; here I can only gesture at such an account. It is well known that the picture of perception suggested by much recent work in the philosophy and psychology of perception (and to my mind, supported by the existence of the Waterfall Illusion) is that of the perceptual system as an *information processor.* The system is in states with informational content, and they normally produce beliefs which have those very same contents. This idea is quite innocuous: the perception that *p* normally produces the belief that *p*.

When we have beliefs, their contents are composed of concepts; we possess (whatever this might mean) the concepts which are the constituents of the contents. It may appear that the perception that *p* is composed of the concepts that the belief that *p* is. But this is a mistake, resulting from assimilating the nature of belief to the nature of perception. On the picture I favour, the possession of concepts is part of what marks the distinction between someone who has genuine beliefs (and other propositional attitudes) and a mere information processor. The contents of the states of the perceptual system are, of course, *describable* (by a theorist) in terms of certain concepts, but the system does not *possess* those concepts. It is not a thinker or an agent, and has no propositional attitudes. It is in states with content in the way that a thermometer is: a thermometer can indicate that the temperature is 70 degrees without possessing the concept of a degree, or of temperature. There is no general problem about things being in such states — the perceptual system is just another instance. As I see it, one of the attractions for the philosophy of mind of the idea of an information processing theory of a perceptual system (e.g. the auditory or visual systems) is the possibility of an explanation of how perceivers can come to be in states of a certain kind in virtue of their perceptual systems being states of other kinds. (A simple analogy might help: our lungs expand and contract, and because they do so, we breathe. But the *lungs* do not breathe.)

However, it may be responded that although *perceptual systems* are not possessors of concepts, *perceivers* do need to possess (perhaps tacitly) the concepts which can be used to characterize the informational output of the perceptual systems, in order to have beliefs based on perception. I do not agree. For one thing, this response rests on a dubious notion of concept possession, which entirely obscures the well-established distinction between personal and sub-personal psychological states. For another, it is *prima facie* implausible to suppose that perceivers need to possess the concepts which *can be used* to characterize the content of perceptual states, in order to be in those states. Of course, someone who knows a lot about (say) the visual system will possess these concepts, but he will not need to possess them in order to see. Information-processing theories of vision (such as David Marr's (1982)) attribute contents to states of the visual system which are extremely complex; it is implausible to suppose that a perceiver must be a master of the concepts involved in these contents in order for his visual system to be in those states (for a discussion of this sort of point, see Davies 1986, p. 144). In my view, it is more plausible to suppose that belief formation *conceptualizes* the content of perceptual states. In the Waterfall case, the output of the visual system is a contradiction, perhaps because (a) of a failure of the mechanism (perhaps in the way Mellor describes); and (b) because the states of the visual system are 'informationally encapsulated' in Fodor's sense. (This answers Mellor's query (p. 147) about what entitles me to talk about contradictions at all; the contradiction — one which a believer would naturally describe as *Fa* & ~ *Fa*, though the information-processing theory would give a more complex description — is actually *in* the informational output of the visual system.) It seems to me that this picture of the structure of perception suggests the possibility of a clearer explanation than belief-theories give of why the contents of perceptual states cannot be revised in response to further evidence. But much more must be said about concept possession before it can be universally accepted that the Waterfall Illusion is evidence for the accuracy of this picture.

Peterhouse
Cambridge CB2 1RD

© TIM CRANE 1988

REFERENCES

M. Davies 1986, 'Tacit Knowledge and the Structure of Thought and Language' in C. Travis (ed.) *Meaning and Interpretation*, Oxford: Blackwell

D. Marr 1982, *Vision*, San Francisco: Freeman and Co

THE 'BY' WORD

By MICHAEL WREEN

MY thanks to Jennifer Hornsby (ANALYSIS 47.4, October 1987, pp. 238–9), Alan White (ibid., pp. 239–40), and Karl Pfeifer (ANALYSIS 48.2, March 1988, pp. 107–9) for their comments on and criticism of my earlier piece on the by-relation (ANALYSIS 47.2, March 1987, pp. 120–4). I remain convinced, however, that the approach I advocate there is fundamentally correct.

The gist of Hornsby's reply is (a) that I criticized her for offering a defective solution to (call it) problem B, the familiar 'by' problem, while what she was really doing, in the section of her book I was discussing, was offering a solution to a different problem, problem A, and (b) that my solution to problem B will not apply, *mutatis mutandis*, to problem A; hence, my criticism of her is doubly defective, and her transitive/intransitive verb solution to problem A remains intact. To fill in the details a bit here: Problem B is that we are tempted to subscribe to both (1) 'John clenched his fist by contracting his muscles' and (2) 'John contracted his muscles by clenching his fist', though subscribing to both would seem to necessitate embracing backward causation. Problem A is that 'the description [of the case] (quoted by Wreen [from Hornsby's book]), though it comes naturally, is absurd' (Hornsby, p. 238). More details than that about problem A are not given, but are recoverable from her book, at least given her claim here. Such details can also be found, implicitly, two paragraphs hence.

In relation to charge (a), I must say that I apologize if I did indeed cross wires. I have to add, though, that even on reading over her second chapter several more times, I see little evidence of an attempt to clearly distinguish two separate problems. Moreover, the very markers that lay beside the passage I quoted, the markers that say 'R. Taylor' and 'von Wright', are misleading in this regard, for as I read the two (and especially Taylor), they point in the direction of problem B, not problem A.

More importantly, if the passage I quoted, or at least some of the sentences in that passage, aren't taken as a colloquial, everyday-discourse way of expressing or getting at problem B, then, Hornsby notwithstanding, the description seems anything but natural, and therefore anything but puzzle generating. It's not natural, not for me or anyone I know, to say, and literally mean, that my clenching my fist causes the muscles in my arm to contract. Absurd, yes, just as Hornsby says; but natural, no, not at all. But then if there's no intuitive appeal to the propositions that supposedly generate a problem, there's no reason to think there's really any problem about at all. That's one reason — but only one (see the next paragraph, too) — why I'm inclined to think that there really isn't any problem A, or at least no problem A distinct from problem B.

Hornsby has also objected, in correspondence, that her problem, problem A, isn't dependent on the use of the 'by' locution, while mine apparently is. Well, I don't really think that the usual problem, problem B, has to involve the use of the word 'by'. Circumlocution or alternative phrasing is certainly possible (see below). But more importantly, I don't think the passage in her book that she refers to in her reply (p. 28 of *Actions*), a passage which supposedly illustrates that point, will help sustain charge (a), that there are two distinct problems, and Wreen has missed the boat on one of them. 'He clenches his fist', she says there, 'and his doing so makes his forearm muscles contract'. No 'by' here. 'Make' certainly sounds causal, though, and the claim thus just as weird, if taken literally, as those in the passage I originally quoted. What Hornsby says after this, however, is decidedly more appealing. What we have in this situation, she says, is 'an action, of clenching the fist, and something else, the muscles' contraction, that is made to happen *in* acting'. (My emphasis. See also *Actions*, p. 24 on 'resulting in' and 'resulting from'.) It's certainly possible to read 'made to happen in' non-causally here, or perhaps causally, but in a way that avoids the causal implication of 'make' alone — though I think that a bare 'makes' in the previous sentence would then be at least misleading. But then — and here I explicitly begin my reply to charge (b), that my solution to problem B will not work, *mutatis mutandis*, for problem A — it looks as if what's really on deck is the indirect achievement or performance of an action, or at least a close cousin of such — a close cousin if, as Hornsby and others think, a person's muscles' contracting isn't identical with his contracting his muscles. Thus the problem is, or is very close to, the more usual 'by' problem, and my solution to it, or at least a small extension of such, becomes available, and will suffice here as well. For even on the assumption that a person's contracting his muscles isn't identical with his muscles' contracting, my 'Two "by" ways' solution could be used to link the actions of fist clenching and muscle contracting through the concept of indirect achievement, and then secondly, since a person's contracting his muscles is logically sufficient for his muscles' contracting, the former action linked to the further achievement of a non-action, his muscles' contracting. Thus even generously reading 'makes' as 'makes to happen in', and even granting the non-identity assumption just mentioned, no distinct problem, problem A, arises, and no fundamentally new solution is called for. Certainly theses about transitive and intransitive verbs and their relation to actions and non-actions aren't required.

Even if all of the above is incorrect, however, even if there really is a distinct problem A, even if I inappropriately criticized Hornsby for inadequately responding to B, while all the while she was really discussing A, and even if Hornsby's transitive/intransitive verb distinction is the key to problem A — even if all that's so, my main purpose in 'Two "By" Ways' was to supply a solution to problem B. My solution is independent of all the 'errors' just noted. I can, then, end my remarks on her reply on at least a semi-positive note, and commend reflection on my solution and further reflection on whether, as Hornsby says, there is (implicit but) effective criticism of it in her book, and a better solution 'drawn together' at p. 93.

It should be evident, from what was said two paragraphs back, that White and I have something in common, but also evident that we have our differences. As I read him, he both criticizes my solution and offers an alternative solution of his own, one that he thinks is adequate and doesn't require that 'by' be read ambiguously. Two related objections are advanced against my ambiguous-'by' solution. Consider, he says,

(3) The kinetic energy of the molecules of a liquid alters by the liquid's boiling,

and

(4) The liquid boils by an alteration in the kinetic energy of its molecules.

'The conceptual problem of the role of "by" here is not related to the notion of action at all', he says — that's objection number one — and, moreover, the problem is not 'solved by suggesting that by boiling...a liquid we "indirectly or *a fortiori* achieve or perform" the task of altering the kinetic energy of the molecules' — that's objection number two (White, pp. 239–40).

I grant that the conceptual role of 'by' in such cases isn't actional at all, and thus that appealing to causal generation and indirect action achievement is beside the point. But, first, my original solution was intended only to handle the sort of problem case that has worried Hornsby and others, and second and more importantly, I don't really see any major problems for my solution in any case. For since, as White says, (3) can't be read causally, with the boiling causing an alternation in kinetic energy, the most natural way to read it — and it does seem to me a bit of a strained statement to begin with — is as a remark in an overall explanation of a system of phenomena, or as a statement in a recipe of sorts. As a statement in a recipe, what seems to be indicated is (perhaps White notwithstanding — see the quoted remark) directions for indirect achievement. As a statement in an overall explanation, what is being said would be better put, at least to an American ear, in terms of 'alters with', since the general idea seems to be that one set of phenomenon, alteration of kinetic energy, is caught up with another set of phenomenon, boiling, in the causal network being explained: if one changes (boiling occurs), the other does (there is an increase, or alternation, of kinetic energy); that's why 'alters with', rather than 'alters by', seems more appropriate. The exact nature of the causal network is not indicated in the statement itself. Those, at any rate, are the only two interpretations of (3) that I can think of that allow it to be true. (4) isn't problematic, as the 'by' in it can be glossed causally or in a relatively simple explanatory sense, with 'by' meaning 'explained by'. Thus, even if the context isn't actional, an extension of my solution to the usual 'by' problem will work here, I think. 'By' is multiply ambiguous, with context indicating the sense needed, and the 'by'-locution is available both within and without the actional context. — This, incidentally, also mutes one of Pfeifer's objections, namely that the sort of paradox I address in the earlier paper is 'inherent in "by"-locutionary talk generally and not just the particular case' discussed in the original paper (Pfeifer, p. 107).

White's positive proposal, that 'by' be read univocally, as introducing a sufficient condition of a causal, logical, or some other sort, I find attractive and intriguing. Interpreted one way, though, it introduces just the sort of ambiguity that White seems to want to avoid, since the kind of sufficient condition conceptually indicated will vary from sentence to sentence — varying from extra-linguistic context to context won't do, for reasons that will become clear in a moment — and ambiguity is thus given entrance at the back door. Thus interpreted, in fact, his theory differs but little from mine. However, if 'by' is read in different sentences without the causal, logical, or whatever force indicated, if it is read as a bare introducer of sufficiency, my enthusiasm wanes, as I think it subject to counterexamples, and subject to counterexamples because not sufficiently detailed in its reading of 'by'. Let me explain with a fleshed-out variation on one of White's own examples. Consider

(5) Weber drew an equilateral triangle by drawing an equiangular triangle,

and

(6) Weber drew an equiangular triangle by drawing an equilateral triangle.

'Drawing an equilateral triangle' is logically necessary and sufficient for 'drawing an equiangular triangle', but there may be situations in which we would subscribe to (6) but not (5), or at least subscribe to each for fundamentally different reasons. Suppose that young Weber knew what an equilateral triangle was and how to draw one, didn't know what an equiangular triangle was, but did know that drawing an equilateral was sufficient for drawing an equiangular ('whatever that might be', he thinks), just as it is for drawing a triangle. If the lad were asked to draw an equiangular triangle and complied, (6) would be true, it seems to me, but (5) false, or at best true for fundamentally different reasons. He didn't draw an equilateral one *by* drawing an equiangular one, I'm tempted to say; rather, the converse, and the converse only. On White's analysis, however, it would seem that since (6) is true, (5) would be as well, 'drawing an equilateral triangle' being logically necessary and sufficient for 'drawing an equiangular triangle'. Or if (5) is regarded as true, that would be because the statement occurred in an explanation, given by a third party, in which one phenomenon is understood, but some explanation of its occurrence wanted. The 'by', in short, would have to be read as the 'by' of explanation, and not as the 'by' of logical sufficiency. What's missing and makes such counterexamples, or at least anomalous asymmetries, possible is a detailed reading of the 'by' locution, a reading which may differ from sentence to sentence, and not just non-linguistic context to context.

My treatment of (5) and (6) above applies as well, *mutatis mutandis*, to Pfeifer's

(7) John moves a wooden counter by moving a black counter,

and

(8) John moves a black counter by moving a wooden counter,

with which of the two being true and which false, under the most natural reading of both, depending on John's action plan. Of course we could abandon all pretence of following John's action plan, as Pfeifer does, and read both, and apparently every other 'by', univocally, in terms of a 'by' of explanation. But such a reading does no justice to everyday speech or the concerns of people who utter such sentences, or to our pre-theoretical intuitions about such cases. Special, contrived contexts have to be concocted to make the explanatory 'by' apropos.

Last, Pfeifer objects that indirect achievement is itself a form of action generation on Goldman's account, namely simple generation, and thus the distinction I wish to draw between act generation and indirect achievement is muted, or even nonexistent. To this I say, So much the worse for Goldman's concept of simple generation. The fact that I gestured in a friendly way toward Goldman in regard to causal generation doesn't mean that I'm committed to accepting everything he says. And it should be obvious from my treatment of cases considered above that I don't in fact accept his concept of simple generation.

Still, I think that Pfeifer is – indirectly, if I may be permitted to say as much – on to a good point here, namely that I don't really define indirect achievement or performance in the earlier paper, but simply rely on an intuitive understanding of the concept. The problem is thus one of deficiency: a fully satisfying resolution of the paradox requires that such a central explanatory notion be clarified or explicated, if not defined outright. To that charge I must currently plead guilty, but beg the mercy of the court for another occasion on which to clear the docket.

Marquette University,
Milwaukee, WI 53233, U.S.A.

© MICHAEL WREEN 1988

ON A THEORY ABOUT THEORIES ABOUT ALL THEORIES

By GEORGE REA

IN 'Tu Quoque' (ANALYSIS 47.4, October 1987, pp. 199-201), Roger Teichmann fails to find a satisfying formulation of the following principle.

> (P) Any theory laying down normative conditions for theories in general should itself meet those conditions.

(P) seems to be true, and as Teichmann indicates, the reasoning behind (P) can be supported in terms of a reductio argument. But on a superficial inspection it can appear that (P) lays down normative conditions for theories in general and so, by its own principle, should meet its own conditions. It seems to be unable to do this because its own conditions seem to be namely meeting its own conditions, namely meeting its own conditions...This superficial analysis is a mistake. The main thing we need to note is that (P) is not a theory which lays down conditions for theories in general. (P) lays down conditions for only a particular sort of theory, viz. a theory which lays down normative conditions for theories in general, (and (P) is not a theory of this particular sort). This can be better appreciated if (P) is expressed more clearly. The sort of theory (P) is concerned with is a theory which claims that certain conditions must be met by any theory if it is to be adequate — for instance conditions which must be met by a theory if it is to be meaningful. This is the significance of 'conditions for theories *in general*'; conditions which must be met by *any* adequate theory. A better way of expressing (P) which clarifies the fact that it is theory about other sorts of theory, and that it just asserts that they must satisfy *their* conditions, goes as follows.

> (Q) *Any theory which lays down conditions for ALL theories* should itself meet those conditions.

(Q) is not a theory of the type described in italics — which I shall call a universal theory. In other words (Q) is not a theory which lays down conditions to be met by *all* theories. (Q) merely lays down conditions which are to be met by a *particular type* of theory (viz. a universal theory, which (Q) is not). Of course, even though (Q) does not come under the scope of its own theory, it does come under the scope of universal theories (which apply to all theories), so (Q) will be expected to satisfy *their* conditions if they are correct. But we have no reason to suppose that (Q) will fail to do this. For instance (Q), as we have clarified it, does seem to be meaningful, and if so it will satisfy any adequate theory which lays down certain conditions required by any theory if it is to be meaningful.

8 Highfield Road,
Cleator Moor, Cumbria

© GEORGE REA 1988

I'M GLAD YOU ASKED ME THAT QUESTION

By LAWRENCE ZALCMAN

'HAVE you ever noticed that, while most questions aren't their own answers, some are?'

'How's that?'

'Well, questions like "Who is Sylvia?" or "What is truth?" or "When shall we three meet again?" or "Where are the snows of yesteryear?" or "How long, Catiline, will you abuse our patience?" or "Is this a dagger which I see before me?" aren't their own answers, are they?'

'Isn't that obvious?'

'And isn't it equally obvious that *other* questions such as "What did I just say?" or "What is an example of an interrogative sentence in English?" or (*holding up a copy of Raymond Smullyan's puzzle book,* "What is the name of this book?") "What is the name of this book?" *are* their own answers?'

'What are you getting at?'

'Just this: Is the question "What is an example of a question which is not its own answer?" its own answer?'

'Now why didn't I think of that?'

Dept. of Mathematics
Bar-Ilan University
52100 Ramat-Gan
Israel

© LAWRENCE ZALCMAN 1988

OXFORD

Mind *The Journal of the Mind Association*

Editor: **Simon Blackburn,** Pembroke College, Oxford

Mind is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy.

Forthcoming and recent articles include:

The Narrow Semantics of Names **Michael McDermott**

Personal and Impersonal Identity **T L S Sprigge**

Beyond Skepticism, to the Best of Our Knowledge **Ernest Sosa**

The Difference between Feeling and Thinking **Stephen Everson**

Can Truth be Relativized to Kinds of Mind? **Leslie Stevenson**

Subscription rates for Volume 97, 1988 (4 issues):

Institutional rate: UK £18, N America US$38, Elsewhere £22
Individual rate: UK£11, N America US$25, Elsewhere £15

To order, please apply (with correct remittance) to the Journals Subscription Department at address below, or send for a **free sample copy**. Please note that personal subscriptions are sent to private addresses only. Payment may be made by cheque or major credit card.

Oxford University Press, Walton Street, Oxford OX2 6DP, UK

JOURNALS

---

the review of

metaphysics

a philosophical quarterly

MARCH 1988 | VOL. XLI, No. 3 | ISSUE No. 163 | $9.00

articles

STEPHEN WATSON *Heidegger, Rationality, and the Critique of Judgment*

MICHAEL MURRAY *Husserl and Heidegger: Contructing and Deconstructing Greek Philosophy*

JOHN CAPUTO *Demythologizing Heidegger:* Aletheia *and the History of Being*

MICHAEL WILLIAMS *Scepticism without Theory*

critical studies

DANIEL DAHLSTROM *Heidegger's Last Word*

ALISTAIR MORRISON and Staff *Summaries and Comments*

philosophical abstracts *announcements*

Individual Subscriptions $23.00 Institutional Subsc.iption $35.00 Student/Retired Subscriptions $10.00

Jude P. Dougherty, Editor
The Catholic University of America, Washington, D.C. 20064

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, Jane Heal, R. F. Holland, C. Hookway, C. A. Kirwan, D. H. Mellor, H. W. Noonan, M. A. Stewart, A. R. White, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England, and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages; appearing in January, March, June and October. Orders should be sent to the publisher. Subscription prices for 1988 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£13.00	£16.00	U.S. $33.00
Individuals	£8.00	£10.00	U.S. $21.00

Single Issue Prices

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£4.30	£5.10	U.S. $10.60
Individuals	£3.10	£3.70	U.S. $6.70

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

CONTRIBUTIONS. Articles submitted for publication before 1 September 1988 should be sent to the Editor, Peter Smith, Department of Philosophy, University College of Wales, Aberystwyth, Dyfed SY23 3DY. After that date, the Editor's address will be Department of Philosophy, University of Sheffield, Sheffield S10 2TN.

Contributors are asked to send two copies of their articles, which should normally not exceed 3,000 words in length. Footnotes should be kept to a minimum, and wherever possible be avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them, and the spacing of symbols within formulae should be carefully attended to. Single quotation marks are to be used except for purposes of internal quotation and "scare" quotes. The *Oxford Dictionary for Writers and Editors* is the preferred authority on questions of spelling, etc.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

5 APR 1989

Vol. 48 No. 4 (New Series No. 220) October 19

ANALYSIS

Edited by
Peter Smith

CONTENTS

Internalism and coherentism: a dilemma Paul K. Mos
Reply to Moser Laurence BonJo
How not to be a coherentist Paul K. Mos
The ins and outs of 'metaknowledge' John Bend
A new and unimproved version of reliabilism Casey Swa
Foundationalism, conceptual regress, and reliabilism Richard A. Fumert
The principle of disbelief Leopold Stubenbe
Putnam's miraculous argument Gary Iseming
Realism without truth Rod Bertol
Realism without truth: a response to Bertolet Michael Dev
Closed time and causal loops: a defence against Mellor Susan We
Permissions and illocutionary act taxonomy Christopher Ne
Morally acceptable divestiture Kerry S. Walte
The contingent *a priori*: a reply Timothy Williamso
Reply to Lowe on ships and structures Harold W. Noon
Why only the best is good enough Stephen Grov

BASIL BLACKWELL · OXFORD AND NEW YORK

ISSN 0003-2638

ANALYSIS 50

CALL FOR PAPERS

A Conference to celebrate the 50th Volume of ANALYSIS will be held at Newnham College, Cambridge, March 30 to April 1, 1990. Submission of papers in the ANALYSIS style of not more than 3000 words is invited. Selected papers will be presented at the Conference and published in an expanded March 1990 issue of the journal.

Papers to be considered for the Conference must reach the Editor, Department of Philosophy, The University of Sheffield, Sheffield S10 2TN, by June 30, 1989, and should be clearly marked 'ANALYSIS 50'.

Although there is no restriction on the topics which may be discussed, it is hoped that some contributions will pursue familiar arguments and themes first broached in ANALYSIS.

Further details of the conference will be advertised later.

ANALYSIS 48.4 OCTOBER 1988

INTERNALISM AND COHERENTISM: A DILEMMA

By Paul K. Moser

IN *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press, 1985), Laurence BonJour has argued that only an internalist coherence theory of justification can solve the epistemic regress problem, the problem of explaining how one empirical belief can be justified on the basis of another empirical belief. BonJour's theory is internalist given its implication that one must have some sort of *access* to the justifiers of one's justified beliefs (pp. 10, 23, 31); and it is coherentist given its implication that every justified empirical belief is justified by coherence with the set of one's other beliefs (p. 92). This paper shows that BonJour's combination of internalism and coherentism generates a fatal dilemma.

BonJour's coherentism relies on the familiar coherentist assumption that the basic source of empirical justification is one's *entire system* of beliefs. On this assumption, a particular empirical belief derives its epistemic justification from its cohering with (i.e., sustaining an appropriate inferential relation, such as an explanatory or implication relation, to) one's overall belief-system. Thus, BonJour holds that 'the primary justificatory issue is whether or not, under the presumption that I do indeed hold approximately the system of beliefs which I believe myself to hold, those beliefs are justified' (p. 103).

However, given its internalism, BonJour's coherentism must also specify the sort of required access to one's overall belief-system. A troublesome question is whether this required access is 'cognitive' in BonJour's sense; that is, whether it essentially involves one's *judging* that something is the case, such as that certain beliefs constitute one's overall belief-system. If, on the one hand, the access is not cognitive in this sense, then, given BonJour's coherentism (and its rejection of the traditional doctrine of givenness), it cannot provide a source of justification (p. 69); nor, given BonJour's version of internalism, can it provide the needed possession of, or access to, a justifying reason for a justified belief. According to BonJour's internalism, an empirical belief, *P*, is epistemically justified for a person, *S*, only if *S* is in *cognitive* possession of a reason making *P* likely to be true; and 'the only way to be in cognitive possession of such a reason is to believe *with justification* the premisses from which it follows that the belief [that *P*] is likely to be true' (p. 32;

cf. pp. 43, 123). Thus, BonJour claims that 'if a particular belief *B* is to be justified for a particular person by virtue of possessing [a certain] property, then (if externalism is to be avoided) that person must believe with justification that *B* does in fact have this property' (p. 80; cf. pp. 31, 50). Such internalism underlies BonJour's rejection of foundationalist solutions to the regress problem, according to which a foundational empirical belief can be justified independently of the justification of any other belief. The rejection of such internalism leaves BonJour without a challenge to foundationalism.

If, on the other hand, the required access to one's overall belief-system is cognitive in BonJour's sense, an equally serious problem arises. In keeping with BonJour's internalism, suppose the cognitive access relevant to my justified belief that *P* is a further belief, specifically the belief — call it '*A1*' — that the members of my overall belief-system are *B1*, *B2*, . . . , *Bn*, and that *P* coheres with those members. Given BonJour's internalism, *A1* by itself cannot provide the needed cognitive possession of a justificatory argument for *P*. For that purpose, according to BonJour, *A1* must itself be justified, since otherwise it would not be a preventive against 'epistemic irresponsibility' by preserving the essential connection between justified belief and high likelihood of truth (cf. pp. 31, 43, 80, 123). But this means that there must also be a justificatory argument supporting *A1*, an argument whose premisses I must believe, according to BonJour's internalism. Call the latter belief (that the relevant premisses are true) '*A2*'. Just as *A1* could not provide the needed cognitive possession of a justificatory argument for *P*, so also *A2*, on BonJour's internalism, cannot provide by itself the needed cognitive possession of a justificatory argument for *A1*. For the latter purpose, *A2* must itself be justified. But this means that we need an *A3* in support of *A2*, and an *A4* in support of *A3*, and so on *ad infinitum*.

Thus, BonJour's internalism leaves us with an infinite regress of required justified beliefs, an endless regress of the following form:

A1: My belief-set is *B*, and *P* coheres with *B*,
A2: My belief-set is *B*, and *A1* coheres with *B*,
A3: My belief-set is *B*, and *A2* coheres with *B*,...

This regress *is* endless because at no point do we arrive at a member that is already justified by an earlier member in the series. Each member in the series is a logically distinct belief; and, on BonJour's internalism, each such member requires a logically distinct justifying *and justified* belief concerning its coherence with one's overall belief-system.

However, it is implausible to suppose that the justification of every empirical belief requires an infinity of additional justified beliefs. A justified belief concerning a proposition's coherence with one's overall belief-set is not a mere *disposition to believe*; it is an actual, distinct belief-state. Presumably, the formation of a distinct belief-state concerning a proposition's coherence with a belief-set takes time, given that a belief-state requires assenting, which, of course, takes time. If this is so, we have good reason to deny that every justified empirical belief requires an infinity of additional justified beliefs. And, even aside from this point, there is, contrary to BonJour's internalism, no reason to think that any of us has the sort of endless series of justified beliefs noted above. The members of such a series involve increasingly complex, iterated levels of coherence assessment, and thus require a degree of conceptual sophistication that goes beyond the requirements of simple, non-iterative justification.

One might try to avoid the present threat of an endless regress by simply stipulating that the belief providing cognitive access to one's overall belief-system is not in need of justification. In fact BonJour evidently favours this move (pp. 106, 147). However, it is *ad hoc* and troublesome. It is *ad hoc*, because its sole purpose is to save BonJour's coherentism from the aforementioned infinite regress problem generated by its internalism. And it is troublesome for two main reasons. First, it leaves us with the mystery of how justified belief can derive ultimately from *un*justified belief, thereby raising the question whether BonJour's coherentism preserves the essential connection between justified belief and high likelihood of truth. And second, it conflicts with BonJour's anti-foundationalist internalist stricture that a judgemental cognitive state requires justification if it is to play a role in epistemic justification. In relaxing that stricture for his own account, BonJour undercuts the main motivation for his coherentism, given that he thereby removes the basis for his rejection of foundationalist solutions to the epistemic regress problem.

We have, then, a straightforward dilemma for internalist coherentism as developed by BonJour: either its internalism generates an implausible endless regress of required justified beliefs, or it loses its motivation as a coherentist alternative to foundationalism. Both horns of this dilemma are fatally sharp: they imply that BonJour's internalist coherentism is either implausible or unmotivated. The general positive lesson of this dilemma is threefold. First, coherentists should drop BonJour's internalist notion of cognitive access in their effort to challenge foundationalist solutions to the epistemic regress problem. Second, internalists need to identify a type of cognitive access to justifiers that does not itself consist in justified belief. Third, and finally, we still need a viable solution to the notorious epistemic regress problem that has troubled epistemology since the time of Aristotle.

Loyola University of Chicago
6525 North Sheridan Road,
Chicago, IL 60626, U.S.A.

© PAUL K. MOSER 1988

REPLY TO MOSER

By LAURENCE BONJOUR

PAUL K. MOSER (ANALYSIS above, pp. 161–3) argues that the combination of internalism and coherentism represented by my account of empirical justification in *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press, 1985 — hereafter *SEK*) 'generates a fatal dilemma'. My response is that the supposed dilemma and the quite genuine problem underlying it were already dealt with explicitly and at length in *SEK*. The solution offered there may or may not be finally acceptable, but since Moser neither explicitly objects to nor indeed even takes any clear notice of it in his paper, I can do little here but reiterate the essential points, referring the reader to *SEK* for more details.

The underlying problem, of course, concerns the epistemic status of my representation of my own system of beliefs in the context of a coherence theory of empirical justification. Clearly the metabelief that I have such-and-such a system of beliefs is contingent and thus seemingly in need of empirical justification. But for a coherence theory, all empirical justification depends on the fact that the justificandum belief coheres with an overall system of beliefs, where for any plausible version of such a theory, the system in question must be the one that the believer actually holds. The inescapable consequence is that such a metabelief (or system of metabeliefs) cannot be justified along coherentist lines — not, primarily at least, because of the regress which Moser constructs, but simply because any such justification would be hopelessly and unavoidably circular, dependent on the acceptability of the very representation of my system of beliefs whose justification is at issue (*SEK*, p. 102). And this result threatens to undermine the whole project of a coherentist account of empirical justification.

The response to this problem which is suggested in *SEK* centres around what is there referred to as *the Doxastic Presumption*. Reduced to its barest essentials, the suggestion is that for a coherence theory, all empirical justification must be understood as relativized to or conditional on the presumption that my representation of my system of beliefs is at least approximately correct (only approximately correct because the issue of whether I have some particular belief or relatively small set of beliefs, unlike that pertaining to my whole system of beliefs, is capable of being settled on coherentist grounds (p. 104)). If *P* is the claim whose empirical justification is at issue, then what a successful coherentist justification shows is that *if* this presumption is correct, then *P* is likely to be true (pp. 104–106, 147).

This sort of solution attempts in effect to go between the horns of Moser's dilemma, though not without paying a price. The obvious question which it raises is why the need to appeal in this way to the Doxastic Presumption does not show that the coherentist justification in question is simply inadequate because it fails to show that the belief in question is genuinely likely to be true. The answer I would offer is that this result follows only if adequacy requires that likelihood of truth be established in a way which requires no background assumptions of any sort (or, which comes to the same thing, in a way which can meet any possible sceptical challenge). Unfortunately, however, as argued in *SEK* (pp. 13–15), any imaginable epistemological view is inadequate when judged by such a standard. The interesting question, I suggest, is whether there is any alternative epistemological position, similar in structure at least to foundationalism, which can get by with a weaker presumption and which is otherwise at least approximately as adequate as the coherentist view in question. I have argued in *SEK* (pp. 79–84) that there is not, but that argument is admittedly less than fully conclusive, and it is always possible that further positions will be forthcoming.

I will conclude with a brief look at Moser's threefold 'general positive lesson'. I have already suggested my response to part three of the "lesson": the solution to the regress problem offered in *SEK* may not be viable, but Moser has not shown it not to be viable because he has not even discussed it. As regards the first part of the "lesson", the suggestion that coherentists should abandon internalism, I will say only that, as also discussed in *SEK* (pp. 101–102), internalism is needed to give coherentism any point: if externalism were acceptable, then a foundationalist version would be vastly more straightforward. The other part of the "lesson" is the suggestion that 'internalists need to identify a type of cognitive access to justifiers that does not itself consist in justified belief', thereby avoiding any need for further justification. Presumably this is a veiled allusion to Moser's own internalist foundationalist account in his book *Empirical Justification* (Dordrecht: Reidel, 1985). Though I cannot undertake a detailed discussion of that account here, let me offer the opinion that a careful study of it will suggest strongly that it achieves a kind of 'access to justifiers' that avoids the need for further justification only by depriving that access of any content which can do justificatory work. Moser's "lesson" thus seems to me spurious on all counts.

University of Washington,
Seattle, WA 98195, U.S.A.

© LAURENCE BONJOUR 1988

HOW NOT TO BE A COHERENTIST

By PAUL K. MOSER

LET us overlook the drab interpretive issue whether BonJour, in *The Structure of Empirical Knowledge*, somehow anticipated my anti-coherentist dilemma. I find that he did not, but this is uninteresting history. The philosophically important issue is whether internalist coherentism of BonJour's stripe can escape my dilemma. I find that it cannot.

Recall the gist of my dilemma. Internalist coherentism states that one's overall belief-system is one's *only* source of justifying reasons, and that one must have *access* to one's justifying reasons. The dilemma's first horn is: If the required access is cognitive in BonJour's sense (i.e., essentially a belief-like judgement) and requires justification, then we get a vicious endless regress of required justified beliefs. The second horn of my dilemma is: If the required access is cognitive but does not itself require justification (or even if it is noncognitive), then BonJour's coherentism loses its main motivation as an alternative to foundationalism. BonJour's key anti-foundationalist stricture (*SEK*, p. 69) is that a belief-like judgement must itself be justified to play an essential role in epistemic justification. BonJour also holds that the access in question is indeed cognitive (pp. 32, 43, 80, 123).

BonJour believes he can slip through the horns of my dilemma with help from his Doxastic Presumption: the presumption that one's representation of one's belief-system is at least approximately correct. On BonJour's view the Doxastic Presumption is unjustified and unjustifi*able* (pp. 106, 147). Can the Doxastic Presumption take us through the dilemma's horns? Surely it cannot. *The second horn of my dilemma stems directly from the Doxastic Presumption.* The penultimate paragraph of 'Internalism and Coherentism' (ANALYSIS above, p. 163) anticipates an appeal to the Doxastic Presumption (see the pages cited), and it specifies why such an appeal is *ad hoc* and troublesome. In fact, that very paragraph erects the second horn of my dilemma. So the Doxastic Presumption impales the coherentist on the dilemma's second horn.

I have suggested that BonJour's use of the Doxastic Presumption fails to preserve the essential connection between justified belief and adequate likelihood of truth. BonJour's reply (ANALYSIS above, p. 165): 'this result follows only if adequacy requires that likelihood of truth be established in a way which requires no background assumptions of any sort . . . [but] any imaginable epistemological view is inadequate when judged by such a standard'. I doubt both parts of this reply. We plausibly can acknowledge an epistemic role for background assumptions *so long as they are justified but do not depend for their justification on other beliefs*. Also I happen to have formulated a theory of evidential probability (I trust it is at least imaginable) that does without unjustified background assumptions such as BonJour's. (See my *Knowledge and Evidence*, forthcoming 1989 from Cambridge University Press.)

BonJour's reply misses the point in any case. He cannot consistently embrace unjustified background assumptions in his coherentism while using his aforementioned anti-foundationalist stricture (*SEK* p. 69) to reject certain versions of foundationalism. If the foundationalist cannot plausibly attribute an essential epistemic role to unjustified beliefs, the coherentist cannot either. Zoology recapitulates epistemology: what's bad for the goose is bad for the gander also.

If sound, BonJour's anti-foundationalist argument leads not to coherentism, but to *scepticism*. This is an additional lesson of my dilemma. But BonJour's argument rests on a false premiss: the premiss that psychological contents 'can do justificatory work' only if they are propositional or judgemental. This is another unjustified assumption of BonJour's coherentism. Such an assumption gives us no genuine hope of escaping Aristotle's regress problem. So we face another dilemma: scepticism or foundationalism. For reasons presented elsewhere, I recommend foundationalism.

Loyola University of Chicago,
Chicago, IL 60626, U.S.A.

© PAUL K. MOSER 1988

THE INS AND OUTS OF 'METAKNOWLEDGE'

By JOHN BENDER

IF Keith Lehrer is a trustworthy source of information about things epistemological, the debate between 'internalist' and 'externalist' theories of knowledge is the result of a false dichotomy, and the correct analysis of knowledge involves both external and internal factors.

Such theoretical arbitration should be valued whenever we can get it, and I have no grievance in principle against a mixed view. Nor is my present concern with the detailed workings of Lehrer's proposal in terms of coherence, a proposal that is, in a sense, a mixed view. My worry, in fact, is that the theory itself does not support the rather new gloss of it that Lehrer offers, and in terms of which he proposes the dichotomy's resolution (in 'Metaknowledge: Undefeated Justification', *Synthese* 74, 1988, pp. 329–47). According to this gloss, knowledge is a meta-mental state requiring correct (i.e. true) evaluations of lower-order information by the knower. This paper investigates this claim, its relation to the coherence theory of knowledge and justification, and the dichotomy between internalism and externalism.

I CONNECTING HIGHER-ORDER EVALUATIONS TO COHERENCE

What must be added to a true belief to yield knowledge, according to Lehrer, is a particular sort of justification which has both an internalist and an externalist component. The internalist element is the evaluation by the subject of the incoming information in terms of already accepted background information. Lehrer variously describes this evaluation as having to do with the trustworthiness or non-deceptive nature of the information, the source or mechanism yielding the information, and with one's own trustworthiness as a believer and evaluator of information of this sort in these kinds of circumstances.

This meta-mental certification connects up to Lehrer's well-known coherence theory of justification in the following way: for *S* to know that *p* she must be (*inter alia*) personally justified in accepting that *p*, and personal justification is a matter of the coherence of *p* with the background system of *S*'s acceptances. To cohere with the acceptance system, *p* must beat or neutralize every statement with which it competes. Lehrer defines these terms as follows:

> *p competes* with *q* for *S* on the basis of system *A* if and only if it is more reasonable for *S* to accept *p* on the assumption that not *q* than on the assumption that *q* on the basis of system *A*.
>
> *p beats q* for *S* on the basis of system *A* if and only if *p* competes with *q* for *S* on the basis of system *A* and it is more reasonable for *S* to accept *p* than to accept *q* on the basis of system *A*.
>
> *n neutralizes q* as a competitor of *p* for *S* on the basis of system *A* if and only if *q* competes with *p* for *S* on the basis of system *A*, and *n* is such that the conjunction of *q* and *n* does not compete with *p* for *S* on the basis of system *A* [and] it is as reasonable for *S* to accept the conjunction of *q* and *n* as to accept *q* alone on the basis of system *A*. ('The Coherence Theory of Knowledge', *Philosophical Topics* 14, 1986, pp. 5–15, at p. 8)

Among competitors to *p* will be sceptical innuendoes such as, 'sometimes people deceive', 'people sometimes dream', 'sometimes circumstances are not trustworthy'. A positive evaluation of the information received, its source, or one's trustworthiness in the present circumstances is used by Lehrer to neutralize such sceptically motivated competitors to *p*, and is the internalist element of his theory.

The externalist feature of justification comes in the form of the requirement that the subject's positive evaluation of the incoming information that *p* must be correct. If I understand Lehrer here, this condition is satisfied within the coherence model by the demand that coherence is maintained when error is removed from the subject's background system. The 'ultrasystem' of the person is defined as the set of systems that includes the person's acceptance system as well as any system that can be constructed from the acceptance system by either deleting one or more accepted falsehoods, or by replacing those falsehoods with their denials. Coherence of a statement, *p*, with every system which is a member of a person's ultrasystem yields the notion of 'undefeated justification'. Since coherence would fail in some member of the ultrasystem when *S*'s evaluation of trustworthiness is false, the conditions for undefeated justification appear to include both the internalist and externalist elements mentioned. Knowledge is equivalent to undefeated justification, according to Lehrer, since *S* knows that *p* at *t* if and only if *S* is justified in accepting that *p* at *t* in a way that is undefeated.

II What is it to Evaluate Trustworthiness?

I want to focus now on the particularly puzzling passage in which Lehrer makes the claim that human knowledge requires correct and positive evaluations of the incoming information, its source, and circumstances. From 'Metaknowledge: Undefeated Justification':

> If a person receives the information that *p*, all that person would need to know in order to know that *p* is that the information he or she received is correct.... Another may tell us the truth, deliver information that *p* to us, but when we have no idea whether the information delivered is correct, whether the person is trustworthy, for example, then we do not know that the information we have received is correct, and, therefore, we do not know that *p*. A similar point concerns our senses, memory and any other faculty or belief forming mechanism. If we do not have any idea whether the information delivered is correct, if we have no idea whether, for example, our eyes are trustworthy, then we do not know whether the information we receive from that sense is correct. It, therefore, falls short of knowledge. To know that *p* when we receive the information that *p*, we must evaluate the incoming information in terms of background information.... [I]f we are trustworthy at evaluating information, as we assume, then we know that the information we receive from our senses and other faculties is correct, and we possess knowledge. (pp. 334, 344–5)

One interpretation of the claim that, when we receive the information that *p*, we know that *p* only if we know that the information that *p* is correct, can be safely set aside as trivial and not germane to Lehrer's concerns. If we are confronted with the information that *p*, or an assertion that *p*, or find ourselves with a tendency to believe that *p* on the basis of current sensory stimulation, we must, if we are to know that *p*, come to believe that *p* is true (correct). If we have no idea whether the information is correct, then we do not accept that *p* and consequently do not know that *p*. This is not Lehrer's point. There would be nothing new in claiming that knowledge that *p* entails believing that *p* is true.

The implication which *is* of concern is the one Lehrer expresses by saying that knowledge is metaknowledge. To know, one must evaluate the information, and come to accept that it is trustworthy. But what, precisely, does this demand for such an evaluation amount to? In the remarks above, Lehrer distinguishes three states: the receiving of information, the evaluating of information for trustworthiness, and the knowledge state. Since knowledge implies acceptance or belief, we may also distinguish among reception of information, evaluation of information, and acceptance of information. Reflecting on these states, Lehrer's concern with scepticism, and the passage cited, I can differentiate three interpretations of the demand that we correctly and positively evaluate the trustworthiness of incoming information if we are to know.

1. The source of the information that *p*, the cognitive mechanisms involved in the reception of the information, and the conditions surrounding the reception of the information that *p* must be correctly judged by *S* to be of a reliable sort.

2. A particular true belief regarding the information that *p* must be part of *S*'s acceptance system, viz., the evaluation that the information that *p* is correct.

3. In coming to accept the received information that *p*, *S* must correctly believe that he or she is not being deceived in some way; i.e., *S* must in some sense infer *p* from the evaluation that the received information that *p* is non-deceptive.

III The Interpretations and Their Problems

1. Lehrer may be requiring that some true, *general* judgement of reliability be made by the knowing subject that is relevant to the received information that *p*. In some cases, this may be a judgement of the trustworthiness of the source of the information, the reliability of the perceptual or cognitive mechanisms involved, or an evaluation that the circumstances surrounding the reception of the information are reliable, trustworthy, or normal. Making such evaluations presumably requires that the subject conceive of the source, mechanism or circumstances as belonging to a certain sort or type which generally leads to true beliefs.

I will only mention and not discuss two obvious objections which Lehrer already has addressed (in 'The Coherence Theory of Knowledge') viz. that we are not aware of having such meta-beliefs in many cases of apparent knowledge, and that his view implies that small children and animals are not knowers in the same sense of the term as you and I are. My objections focus more on how Lehrer's conditions are meant to fit into our conception of justification and knowledge.

Apparently, Lehrer now sees the requirement of the true higher order evaluations of information as part of the justification condition on knowledge, and not as a demand for some fourth, non-justificatory condition. He says that it is a form of justification which answers the Socratic question, 'What must be added to true belief to yield knowledge?'. It follows that, on Lehrer's theory, if a subject's metabelief is false, she not only fails to know that *p*, but also fails to be justified in the sense of that term which is relevant to the analysis of knowledge. But then — if we were living in a demon world in which our perceptual and cognitive mechanisms, inferential and reasoning patterns, canons of evidence, and our beliefs, including metabeliefs about reliability, are all as they are in the actual world, except that the demon renders false everything we accept as true — what we know in the actual world, we would not even be fully justified in accepting in the demon world. The internalist intuition that the two worlds are identical in matters of justification is, I think, extremely powerful. It is difficult to see what defect in the nature or degree of our justification is supposed to explain our lack of knowledge in the demon world. And it is worth pointing out that this conclusion in fact seems to follow from Lehrer's own general understanding of justification in terms of the comparative reasonableness of competing hypotheses. If I *do* know that there is a computer keyboard in front of me now, then it is more reasonable for me to accept this than it is to accept that I am the victim of demonic horseplay. Hence, in the demon world itself, it is more reasonable for me to believe that I see a keyboard than to believe in demons.

However, I do not wish to beg the question whether *some* externalist element of justification *might* be what is lacking in the demon world case. If, for example, it could be argued that we are never justified in believing a proposition whose truth is objectively highly improbable (in a sense of objective probability that would need to be explicated and defended), and if it could be argued that the propositions we believe in the demon world are, in this non-epistemic sense, improbable, then perhaps a "mixed" theory would convince us that we lack justification in the demon example.

But I do wish to question the particular proposal Lehrer has made. Why are we to believe that the *falsity* of a metabelief undermines justification? No one who fears the charge of confusing matters of justification with matters of truth would suggest that the falsity of the *target belief* in the demon case undermines our justification for accepting it, and yet the falsity of the metabelief ultimately amounts to the fact that the preponderance of (in fact, *all*) first-order beliefs of the same sort as the target belief are false in this case. Furthermore, it is unclear why a more natural and modest suggestion from the proponent of the ascent to metabeliefs should not be that justification of the sort that supports knowledge requires a *justified* (but not necessarily true) metabelief concerning trustworthiness. The threat of regress may dampen this proposal, but the ascent to the meta-level itself risks this problem, and moreover, truth cannot be used to stop a justification regress.

Holding false positive beliefs about the trustworthiness or reliability of the sources of, or circumstances surrounding, information may well defeat knowledge without depriving us of justification, in much the way that traditional Gettier examples that essentially involve a false belief fail of knowledge but are cases of complete justification. Moreover, the fact that a false metabelief might undermine knowledge does not imply that a true metabelief is a necessary condition for knowledge. Knowledge may be only negatively conditioned by metabeliefs. For example, *S* may know that *p* only if *S* neither believes nor has good reason to believe that the information that *p* is unreliable in some way. This suggestion may be in need of refinement, but Lehrer has not argued against it. Neither has he provided us with reason for thinking that the standards of trustworthiness we may place on the testimony of others as sources of knowledge will be reflected in similar metarequirements on other sources of knowledge such as sensory perception. Lehrer notes with approval Thomas Reid's claim of similarity between the deliverances of our faculties and the testimony of others. But these sources of information may well be distinct in their degree of reliability, and consequently, even if a strong argument in favour of evaluative metabeliefs about the trustworthiness of others were offered, no conclusion would follow regarding the necessity of metabeliefs in perceptual knowledge.

Nor am I sure that every case of perceptual knowledge is accompanied by an evaluation, or assumption on the part of the subject of his or her trustworthiness regarding information of the present sort, or information delivered in circumstances of the present sort, or the like. Claude is a lover of the paintings of Monet; he has seen and studied many examples in numerous museums, has read and studied about the painter and his times. Claude is sensitive, perceptive, but not an art or painting expert. He realizes that he could not confidently discriminate every work of Monet from the work of other impressionists: there were so many talented followers, and then there are those canvases of Pissarro in the National Gallery that he mistook for Monets.

But now he stands in front of a painting about which there can be no doubt. The visual evidence is overwhelming; everything about the canvas delivers the information, 'Monet'. The fact that some Monets are mistakable for the work of others is admitted but irrelevant here: Claude knows this is a Monet if he knows anything about the painter's work. Does Claude accept or assume that he is trustworthy in his judgments in situations *of this sort*? If the situation is described as one of discriminating Monets from non-Monets, no. But then, under what sort is he conceiving the case? Does his justification and knowledge in the case rely upon or presuppose that Claude has conceived this instance under any epistemically relevant sort or type concept, and judges himself to be reliable about that sort of information? He may view this as a uniquely clear instance. Of course, as a result of believing so strongly that this painting is a Monet, Claude believes that if he were to see other paintings which were equally impressive, he would be able reliably to identify them also, but this hardly amounts to an evaluation of the present situation in terms of Claude's background beliefs about his trustworthiness in similar situations.

2. To say that a source of information is trustworthy is to say something different than that certain information received from the source is true. At most, trustworthiness of the source implies high probability of the information's truth. Nonetheless, as the last remark in the passage quoted above indicates, Lehrer seems comfortable with the idea that trustworthiness implies truth. This creates the second interpretation of the thesis that human knowledge requires evaluations of trustworthiness, and this reading is perhaps most consonant with the assertion that knowing that *p* involves knowing that the information received is correct. What is demanded of the subject, on this interpretation, is an evaluation that the *particular* information he or she has received is true. We seem perilously close here to the trivial reading mentioned above, viz. that knowledge that *p* implies belief that *p* is true. But Lehrer would distinguish an evaluation that the information that *p* is correct from the *acceptance* of *p*. At any rate, I wish to focus on the fact that Lehrer is requiring, on this second interpretation, that the subject evaluate the information that *p* as correct. Justification ultimately is a matter of the coherence of the incoming information with the background information the subject possesses. What matters, Lehrer says, is whether one has background information sufficient to evaluate the information, to know that the information is correct, whatever its origin or sustenance. Coherence is a matter of the comparative reasonableness of certain propositions relative to a background system. Reasonableness is partially determined by probability assessments. It appears that *if* one must evaluate the information that *p* in order to know that *p*, the evaluation suggested by this coherence theory is that it is *highly probable* that the information that *p* is correct. But clearly, this is different from 'knowing that the information received is correct'. Lehrer's coherence theory does not seem to support this latter condition as necessary for knowledge.

In fact, the theory of coherence seems *at best* to yield the even weaker constraint that, given her background system, *it be more reasonable* for *S* to believe that the information received is correct (or that she is trustworthy in these circumstances, or that the source of or mechanisms involved in delivering the information that *p* are reliable, or some such claim) than for *S* to believe any competitor to *p*. And this is different, of course, from the demand that *S* must actually evaluate the incoming information in terms of her background system.

3. Lehrer's long-standing concerns with sceptical challenges and innuendoes may suggest one final reading: if *S* knows that *p* at *t*, then *S* must come to accept that *p* on the basis of a belief that he is not being deceived when he receives the information that *p*. This interpretation fits scenarios of the following sort. Our eyes "deliver" the information that something seems to be red; we evaluate the situation and accept that our senses aren't deceiving us, and consequently accept that there is a red thing in front of us. Or, we receive the information that *p* from Brown: he asserts that *p*. We evaluate as very improbable the possibility that Brown is trying to deceive us, take Brown to be reliable about matters such as *p*, and on these grounds, accept that *p*.

The most plausible construal of these scenarios, however, is in terms of an *inferential* view of justification, and Lehrer has been clear in rejecting such a model. Coherence and inference are distinct. What role is there, then, for beliefs that one's current epistemic circumstances are non-deceptive? Statements to the effect that present circumstances are not deceptive play a central role, for Lehrer, in neutralizing certain kinds of sceptical alternatives that are competitors to propositions we claim to know. If I receive the information that *p* from Brown, the sceptically-motivated statement, 'Sometimes people deceive others', is a competitor of *p*, and must be neutralized. It is neutralized, according to Lehrer, by conjoining it with the statement that Brown is not presently deceiving me, a statement it is reasonable for me to accept.

But now we are back to the point I made against the second interpretation. Must I actually *believe* that Brown is not deceiving me if I am to know that *p*? Or must my background system simply make this statement more reasonable for me than its competition? (Remember that I need not be inferring *p* from a belief that Brown is no deceiver, so the belief isn't necessary, psychologically speaking.) Once again, the charge that we must evaluate the incoming information seems to be conflated with the requirement concerning comparative reasonableness.

The conditions on Lehrer's notion of competition are quite weak, much weaker, for example, than logical contrariness: *p* competes with *c* just in case it is more reasonable for *S* to accept that *p* on the assumption that not-*c* than on the assumption that *c*, given *S*'s background information. Examine a golfer's claim to know that, although his shots may not be the best, if he now swings at this golf ball in front of him, he will at least *hit* the ball. All of the following propositions and many more are competitors of this claim: golfers sometimes lose their balance when swinging; an insect bite can throw off a golf swing; some people get uncontrollable muscle spasms; the heads of golf clubs have been known to fall off during a golfer's backswing; some people have dropped dead in mid-swing. The question is not whether our golfer knows what he claims to know, but whether knowing would require that he have an evaluation of his claim *vis-à-vis* every one of its competitors. This seems unrealistic. But then, neutralization does not imply evaluation.

IV INTERNALISM, EXTERNALISM, AND METAKNOWLEDGE

I have been unsuccessful in finding an interpretation of Lehrer's metaknowledge claim that is both plausible and a consequence of his coherence theory of justification and knowledge. I therefore do not think that the scuffle between internalists and externalists is to be quieted under the banner, 'Human Knowledge is Metaknowledge'. But according to Lehrer, knowledge is metaknowledge is undefeated justification, and perhaps more peace would result if we remove the middleman from this equation and ask if the coherence theory of undefeated justification, *sans* the metaknowledge gloss, answers the debate.

Undefeated justification has both a personal, internalist side (coherence of a given proposition with the subject's acceptance system) as well as an externalist concern with truth and avoidance of error (coherence in the ultrasystem). But as I intimated in offering the first interpretation of Lehrer's position above, coherence to the adjusted systems within the ultrasystem seems to me a fourth, non-justificatory condition on knowledge, an expanded version of the Gettier condition. In other words, a subject can be fully justified in a belief when that belief is not 'ultracoherent'; undefeated justification is not so much a kind or form of justification, but is justification plus the denial that one's justification essentially involves something false. My disagreement with Lehrer on this point is not merely a verbal matter. If the debate between internalists and externalists has been over the nature of the justification condition *narrowly* conceived, if philosophers are internalists or externalists about justification, and not about knowledge, with whatever conditions it entails, then that debate is not addressed or resolved by the suggestion that knowledge is undefeated justification.

And this seems to me to have been the thrust of the controversy. Alston, Goldman, BonJour, and Dancy, to name only a few, each write of the dichotomy in terms of the justification condition. If I am right, then neither Lehrer's metaknowledge gloss, nor his coherence theory of undefeated justification will quiet the internalism–externalism debate.

Ohio University,
Athens, OH 45701-2979, U.S.A.

© John Bender 1988

A NEW AND UNIMPROVED VERSION OF RELIABILISM

By Casey Swank

FOR a belief to be justified is for there to be something (in particular) good about it — but what? It is good for a belief to be reliably produced; for, even if such a belief is not itself true, it is, at least, symptomatic of a reliable, and therefore good, belief-producing process. Is being reliably produced the same thing as being justified, though? Alvin Goldman has recently suggested that the truth is a bit more complicated than this. According to Goldman, for a belief to be justified is for it to be produced in a *presumptively* reliable way — i.e., in a way 'that would be reliable in worlds like the presumptively actual world, that is, in normal worlds' ([2] p. 108). What recommends this revision of reliabilism? And should we join Goldman in embracing the resultant theory?

Among the more dubious implications of (ordinary) reliabilism is one observed by Stewart Cohen. Cohen invites us to consider two believers — '*A*, who is a good reasoner, i.e., reasons in accordance with the canons of inductive inference, and *B*, who engages in confused reasoning, wishful thinking, reliance on emotional attachments, guesswork, etc.' ([1] p. 283). Next, we are to imagine that, unbeknownst to *A* and *B*, their world is controlled by a Cartesian demon, whose machinations render *A*'s reasoning-habits as unreliable as *B*'s. Cohen then observes that, 'Since the beliefs of *A* and *B* are both produced by unreliable processes (the evil demon sees to this), a reliabilist theory of justification must render identical epistemic appraisals of both sets of beliefs' ([1] p. 283) — i.e., assuming reliabilism, *A* is no more justified in his beliefs than *B* is in his.

With Cohen, I think we want to say that, notwithstanding the demon's machinations, 'There is a fundamental epistemic difference between the beliefs of *A* and the beliefs of *B*' ([1] p. 283) — i.e., that, contrary to reliabilism, *A* is more justified in his beliefs than *B* is in his.

And Goldman agrees. Instead of rejecting reliabilism, though, he revises it. Specifically, he suggests that a justified belief is one produced in a *presumptively* (but perhaps not actually) reliable way — i.e., in a way 'that would be reliable in worlds like the presumptively actual world, that is, in normal worlds' ([2] p. 108). Unlike its simpler predecessor, this new version of reliabilism squares with our judgement that, notwithstanding the demon's machinations, *A* is more justified in his beliefs than *B* is in his; for, in a normal world (i.e., a world unplagued by Cartesian demons and the like), *A*'s beliefs would be more reliably produced than *B*'s. There is something, however, with which the new reliabilism does *not* square — viz., the fact that justified is a *good* thing for a belief to be. Let me explain.

If a belief is produced in a presumptively reliable way (i.e., in a way that would be reliable in a normal world), there is presumably something good about it; for the actual world is presumably a normal one, and it is good for a belief to be reliably produced. But this is not yet to say that for a belief to be produced in a presumptively reliable way is, *ipso facto*, for there to be something good about it — i.e., that produced-in-a-presumptively-reliable-way is *itself* a good thing for a belief to be. Is it? I think not. A belief's being produced in a presumptively reliable way is like an airline pilot's being presumptively competent. An airline pilot, competent or incompetent, is no better for being presumptively competent: when we contemplate the possibility that our pilot is in fact incompetent, we are not at all consoled by the realization that he would yet be presumptively competent. Likewise, a belief, reliably or unreliably produced, is no better for being produced in a presumptively reliable way: when we contemplate the possibility that our reasoning-habits are in fact unreliable, we are not at all consoled by the realization that they would yet be presumptively reliable. I would choose the presumptively competent pilot, not because presumptive competence is itself a good thing, but because competence is — and this pilot appears to have it. Similarly, I would choose presumptively reliable reasoning-habits, not because presumptive reliability is itself a good thing, but because reliability is — and these reasoning-habits appear to have it.

Evidently, then, for a belief to be produced in a presumptively reliable way is not (*ipso facto*) for there to be something good about it. Justified is, and produced-in-a-presumptively-reliable-way is not, a good thing for a belief to be. So being thus produced cannot be the same thing as being justified — i.e., Goldman's new reliabilism cannot be correct.[1]

University of Alabama,
Tuscaloosa, Alabama 35487, U.S.A.

© CASEY SWANK 1988

REFERENCES

[1] Stewart Cohen, 'Justification And Truth', *Philosophical Studies* 46 (1984), pp. 279–95.

[2] Alvin Goldman, *Epistemology and Cognition*, Harvard University Press, 1986.

---

[1] For their help with this paper, I wish to thank my wife Mary, and my advisor William Lycan.

FOUNDATIONALISM, CONCEPTUAL REGRESS, AND RELIABILISM

By RICHARD A. FUMERTON

PERHAPS the best known argument for foundationalism claims that it is the only way to avoid an unacceptable vicious regress. But there are at least two importantly different regress arguments for foundationalism and one, a conceptual regress argument, presents difficulties for all but the most primitive reliability analyses of justified belief. Indeed, it provides reasons for returning to a very traditional form of foundationalism.

I THE EPISTEMIC REGRESS ARGUMENT FOR FOUNDATIONALISM

The standard regress argument for foundationalism (I shall call it the *epistemic* regress argument) purports to establish the existence of non-inferentially justified beliefs as the only alternative to the most radical *scepticism* possible. Relying on the principle that in order to be inferentially justified in believing *P* on the basis of *E* one must be justified in believing both *E* and that *E* makes probable *P* (The Principle of Inferential Justification), the foundationalist argues that if all justification were inferential the existence of any justified belief would entail the having of an infinite number of justified beliefs. The only way to end the regress of justification, the argument goes, is to posit the existence of beliefs that are non-inferentially justified.

While the precise analysis of non-inferential justification is a matter of much debate among foundationalists, the majority of *traditional* foundationalists took non-inferentially justified beliefs to be either infallibly or incorrigibly justified. (Let us say that a belief is infallibly justified when its justification precludes the possibility of error. And let us say that belief is incorrigibly justified at a time *t* when its justification could not be *defeated* at *t* through the addition of other evidence.) Such a view, however, *seems* quite unnecessary in order to accept the conclusion of the *epistemic* regress argument. The only obvious conclusion of that argument is that in order to be justified in believing anything, it must be possible to have a justified belief that does not owe its justification to the having of other justified beliefs. One would certainly need a further argument if one wanted to conclude that non-inferentially justified beliefs must be infallibly or incorrigibly justified. Indeed, while traditional foundationalisms have been largely abandoned, contemporary philosophers seem to find much more attractive externalist analyses of justification within which one can often still make a distinction between inferentially and non-inferentially justified beliefs, but where non-inferential justification does not entail either infallibility or incorrigibility as traditionally conceived. Alvin Goldman's now famous reliability theory of justification is a case in point.[1]

II RELIABILITY AND FOUNDATIONALISM

The core thesis of reliabilism is, of course, that justified beliefs are reliably produced beliefs. One gets different versions of reliabilism depending on how one cashes in the crucial concept of reliability. Goldman's reliabilism seems to involve a distinction that closely parallels the traditional foundationalist's distinction between inferentially and non-inferentially justified beliefs. In short, Goldman wants to distinguish conditionally reliable processes which take beliefs as their input, from unconditionally reliable processes which take as their input stimuli other than beliefs. Setting aside some important qualifications that we will discuss later, Goldman identifies an inferentially justified belief with a belief that is produced by a conditionally reliable process where the input beliefs are themselves *justified.* A belief-dependent process is conditionally reliable when the output beliefs are usually true when the input beliefs are true.[2] If this were the only kind of justification recognized by the reliabilist, he would seem to face the same kind of vicious regress that traditional foundationalists sought to avoid. For if the only justified beliefs were beliefs produced by belief-dependent processes whose input beliefs are themselves justified, it is unclear how a finite being could end the causal regress that would seem to be required in order for any belief to be justified. This is no problem for Goldman, of course, since he has the analogue of a foundationally justified belief, namely a belief that is produced by an unconditionally reliable process. Such processes take as their input states other than beliefs, and they are unconditionally reliable when their output beliefs are usually true. Notice that these 'foundationally' justified beliefs need not be either infallible or incorrigible in the traditional sense. The belief-independent process that produces them might not even lawfully guarantee their truth. The process in question might produce output beliefs that are true only ninety per cent of the time. Furthermore, there seems no reason why one might not have a belief produced by a belief-independent unconditionally reliable process while at the same time one could acquire evidence that would defeat the justification provided by that unconditionally reliable process.

---

[1] In what follows I focus on Goldman's now classic paper 'What is Justified Belief?' (in G. Pappas, ed. *Justification and Knowledge,* 1979) rather than his more recent book *Epistemology and Cognition* (1986). I do so because Goldman now seems inclined to abandon the changes suggested in the book at least for what he calls *strong* justification. See his 'Strong and Weak Justification', forthcoming in *Philosophical Perspectives,* Vol. II, 1988. Page references are to the early paper.

[2] I use this oversimplified way of defining the reliability of belief dependent processes for convenience. Goldman would argue, of course, that it is not straightforward frequencies that define reliability. He apparently prefers the philosophically problematic concept of propensities. Nothing in what follows will hinge on the oversimplification I have adopted.

Now when Goldman initially offered his distinction between the two kinds of processes that can result in justified beliefs, he did not do so in the context of avoiding an *epistemic* regress. In fact, you will recall, he proposed a recursive analysis of justification, a recursive analysis which required a base clause. Goldman's primary concern was to provide an analysis of justified belief that did not presuppose an understanding of the concept of justification. One obviously cannot analyse justified beliefs solely in terms of beliefs that are produced by belief-dependent processes that are conditionally reliable where the input beliefs are *justified.* Such an analysis would be blatantly circular. There is, however, the concept of a justified belief that is produced by a belief-independent process that is unconditionally reliable, and one can, Goldman argued, ultimately understand the concept of justification in terms of this base clause. Again, without crucial modifications added later, the idea is that a belief is justified if it is produced by a belief-independent process that is unconditionally reliable OR it is produced by a belief-dependent process whose input beliefs are produced by belief-independent processes that are unconditionally reliable OR whose input beliefs are produced by conditionally reliable belief-dependent processes whose input beliefs are produced by belief-independent processes that are unconditionally reliable OR . . . .

It is useful to reflect on the rhetorical way in which Goldman introduces the analogue of the traditional distinction between inferentially and non-inferentially justified beliefs for it suggests the possibility of a non-epistemic regress argument for foundationalism.

III A CONCEPTUAL REGRESS ARGUMENT FOR FOUNDATIONALISM

One might well take the first task of the epistemologist as that of defining or analysing fundamental epistemic concepts. And if the concept of knowledge presupposes the concept of justification one could argue that the most fundamental epistemic concept that requires analysis is the concept of justified belief. What is it to have a justified belief?

One way to begin answering such a question is to look at paradigms of justified belief most of which, no doubt, will be paradigms of inferentially justified belief. And so we might begin our attempt to understand justification by trying to analyse inferential justification. What is it to be inferentially justified in believing *P* on the basis of *E*? The Principle of Inferential Justification presupposed in the *epistemic* regress argument could be taken as offering a partial analysis of the *concept* of inferential justification. In the context of the epistemic regress argument, its modal status is often ignored, but one might well argue that to be inferentially justified in believing *P* on the basis of *E* just *is* in part to believe *P* as a result of justifiably believing both *E* and that *E* makes likely *P*. As we just noted, however, in examining Goldman's reliability theory, such an analysis of *inferential* justification could hardly be offered as an exhaustive analysis of *justification*. Such a definition of inferential justification clearly presupposes an understanding of the concept of justification.

One might argue, then, that the very concept of *inferential* justification requires that there be some other way of understanding justification. Otherwise one will face a vicious *conceptual* regress. I can understand my inferential justification for believing *P* on the basis of *E* only if I understand my justification for believing *E*. But if all justification were inferential my search for an understanding of what justifies me in believing *E* will take me to yet another justified belief. And if all justification were inferential, I could only understand that justification in terms of still another justified belief, and so on, *ad infinitum*.

An analogy might be helpful. I am suggesting that the foundationalist might argue that understanding inferential justification stands to understanding non-inferential justification, as understanding instrumental goodness stands to understanding intrinsic goodness. Suppose we are told that there is only one concept of goodness, the concept of instrumental goodness. We are then offered the following analysis of instrumental goodness: *X* is instrumentally good when it produces something that is good. It seems intuitively absurd to suggest that we can understand all goodness in terms of instrumental goodness. But why exactly? Notice that one can develop the analogue of our *epistemic* regress argument for foundations. One could argue that if all goodness were instrumental we could never have any reason to believe that anything is good. To know that *X* is good we would have to know that it leads to some *Y* which is good, but to know that *Y* is good we would have to know that it leads to some *Z* which is good, and so on. But there is surely a more fundamental problem with an attempt to identify all goodness with instrumental goodness. Given such an analysis we could never *understand* what it is for something to be good. To understand *X*'s being good I would have to understand something else *Y*'s being good, but to understand *Y*'s being good I would have to understand something else *Z*'s being good, and so on. No matter how hard I try to understand what makes one thing good I am always forced to think of something else's being good. If we assume only that I do understand what it is for something to be good, it follows that I must be able to understand a kind of goodness other than instrumental goodness.

My suggestion is that one can develop an exactly parallel *conceptual* regress argument for foundationalism. One could not even understand inferential justification unless one could understand some other sort of justification upon which the concept of inferential justification is parasitic. Furthermore, I would argue, the *conceptual* regress argument for foundationalism has obvious advantages over the *epistemic* regress argument. For one thing the epistemic regress argument relies on the controversial claim that it is impossible for a finite being to have an infinite number of beliefs. However, if we allow dispositional beliefs, one could surely make a strong argument for the conclusion that all of us *do* have an infinite number of justified beliefs (that 2 is greater than 1, that 3 is greater than 1, and so on). Furthermore, with a conceptual regress argument, one needn't be foiled by the philosopher willing gleefully to embrace scepticism at the expense of foundations. In fact, I believe that *both* the epistemic and the conceptual regress arguments for foundationalism are sound, but my purpose is not to defend that claim in any more detail here. Rather I want to return to Goldman's reliabilism and argue that as necessary qualifications are added to Goldman's initial account, reliabilism runs afoul of the conceptual regress argument for foundationalism. I will conclude only that those who accept the conceptual regress argument for foundationalism had best look to more traditional foundations of *incorrigibly* justified beliefs.

IV Reliabilism and Conceptual Regress

It might seem that Goldman's reliabilism is perfectly compatible with the conclusions of the conceptual regress argument for foundationalism. After all, Goldman's 'inferentially' justified beliefs (beliefs produced by conditionally reliable processes whose input beliefs are justified) are ultimately understood in terms of his 'non-inferentially' justified beliefs (beliefs produced by unconditionally reliable, belief-independent processes). I also noted, however, that Goldman's 'non-inferentially' justified beliefs are neither infallible nor *incorrigible.* And this raises a problem that Goldman himself recognized both in the initial article, his book, and his most recent paper. It seems intuitively obvious that not all beliefs produced by unconditionally reliable processes are justified. Goldman discusses the example of a person whose beliefs about the past are produced by reliable 'memory processes' but who has ample evidence (misleading as it turns out) indicating that these processes are unreliable. Let us suppose that at least some memory processes are belief-independent. We have, most seem to agree, an example of a person whose beliefs about the past are unjustified. In order to resolve this problem Goldman suggests the following revision of his 'base' clause (the clause defining the analogue of non-inferentially justified belief; the clause in terms of which other justification is to be ultimately understood):

> If *S*'s belief in *p* at *t* results from reliable cognitive process, and there is no reliable or conditionally reliable process available to *S* which, had it been used by *S* in addition to the process actually used, would have resulted in *S*'s not believing *p* at *t*, then *S*'s belief in *p* at *t* is justified. (p. 20)

There are obviously a number of highly problematic concepts introduced in the revised base clause, some of which Goldman discusses. But my primary concern is with a necessary part of the new base clause that seems to have been left out. When he introduces the above revision, Goldman tells us that he will 'omit certain details in the interest of clarity' (p. 20). One omitted detail that should certainly concern us is the implicit reference to *justified* belief contained in the reference to available conditionally reliable processes. The availability of a conditionally reliable process that would lead *S* to not believe *p* would only be relevant, I assume, if one had *justified* beliefs that the conditionally reliable process could use as input. The base clause stated more completely and clearly becomes:

> If *S*'s belief in *p* at *t* results from a reliable cognitive process, and there is no belief-independent unconditionally reliable process which had it been used by *S* would have resulted in his not believing *p* and there is no belief-dependent process which is conditionally reliable that could have been used by *S* to process certain *justified* beliefs so as to result in him not believing *p* at *t*, then *S*'s belief in *p* at *t* is justified.

But such a base clause manifestly fails to achieve the purpose of a base clause. In the terminology we have developed, the account of justification offered with such a base clause fails to end a vicious conceptual regress involved in trying to understand the concept of justified belief.

It seems to me that Goldman and reliabilists generally face a dilemma if they accept the conceptual regress argument for the necessity of introducing non-inferentially justified beliefs. Either they return to the crude view according to which the mere fact that a belief is produced by an unconditionally reliable process makes it justified and face obvious counterexamples, or they add the kind of qualifications that Goldman does and run afoul of the conceptual regress argument. Moreover, I think there are more general morals to be drawn. In order to end a conceptual regress involved in trying to understand the concept of justified belief, one may well need something corresponding to the traditional notion of an incorrigibly justified belief. If one is to come up with *non-circular* necessary and *sufficient* conditions for the having of a non-inferentially justified belief, such conditions had better not involve reference to the availability of various sorts of *inferential justification*. The conceptual regress argument, one must remember, supposes that understanding the truth conditions for inferential justification is parasitic upon understanding the truth conditions for non-inferential justification, and not vice versa.

University of Iowa,
Iowa City, IA 52242, U.S.A.

© RICHARD A. FUMERTON 1988

THE PRINCIPLE OF DISBELIEF

By LEOPOLD STUBENBERG

I

IN a paper entitled 'Agnosticism' (ANALYSIS 45.4, October 1985, pp. 219–24) and in chapter 11 of his *Anselmian Explorations: Essays in Philosophical Theology* (Notre Dame, 1987), Thomas V. Morris has defended a principle that 'governs our assessment of positive existence claims in the absence of any grounds for their truth' (Morris 1985, p. 219). The principle specifies a *sufficient condition* for one's being under an obligation to disbelieve a positive existence claim; I shall call it the *Principle of Disbelief* (PD).

> PD For any rational subject S and any positive existence claim P, if (i) S reasonably believes himself to be in good epistemic position relative to P, and (ii) S is in possession of no good evidence or any other positive epistemic ground for thinking that P is true, then S ought to adopt the cognitive relation to P of denial. (Morris 1985, p. 223)

Recently P. J. McGrath (ANALYSIS 47.1, January 1987, pp. 54–7) has criticized PD on three counts. First, PD is too weak because it fails to require disbelief in certain entities in which a rational person would want to disbelieve; second, it has the formal defect of being inapplicable to pure positive existence claims (i.e. unrestricted by spatial and temporal qualifications); and third, it has the absurd consequence of leading to almost total scepticism. He thinks that all three problems are traceable to clause (i) of PD and avers that the correct Principle of Disbelief governing pure existence claims result from purging PD of this clause. I shall argue that McGrath's objections to PD have no force once clause (i) of PD is properly understood.

II

Morris proposes the following definition of 'S is in good epistemic position relative to P':

> Any subject S is in good epistemic position relative to P if and only if (1) P is such that if it were true, there would exist positive epistemic considerations indicating or manifesting its truth, and (2) S is such that if there existed such considerations, he would, or most likely would, possess them. (Morris 1987, p. 207)

Some clarifications are in order. First, positive epistemic considerations (henceforth simply 'evidence') are to be understood as reliable indicators or signs of P's truth; *e* is evidence for P just in case *e* would occur if P were true and *e* would not occur if P were not true. Second, to be in good epistemic position (GEP) relative to P, S must have cognitive capacities suitable for acquiring evidence for P and, third, S must exercise the requisite capacities appropriately in an attempt to acquire the evidence. Keeping these points in mind we can paraphrase Morris's definition of 'S is in GEP relative to P' by saying: *S is in GEP relative to P just in case S would or most likely would have evidence for P if P were true.* Given this understanding of clause (i) of PD I now turn to McGrath's objections to PD.

III

McGrath raises three objections to PD:

> [1] [Morris's] position implies that we should be agnostic about the existence of almost all preternatural beings . . . [such as the Gods of Greece and Rome, tribal deities, elves, hobgoblins, leprechauns, etc.] I am not in good position for assessing their existence . . . a theory which requires us to suspend judgment about the existence of such entities must be open to serious question. (McGrath, p. 54–5)
>
> [2] [There are two kinds of existence claims] — claims that something exists simply (what I shall call 'pure existence claims', e.g. 'God exists') and claims that something exists at a certain time and place (what I shall call 'qualified existence claims', e.g. 'There is a fly in my soup'). The notion of a good epistemic position has application with regard to the second category but not with regard to the first. (McGrath, p. 55)
>
> [3] Morris's position . . . would lead to almost total scepticism about human knowledge. We are clearly in a particularly poor epistemic position with regard to the existence of Descartes' evil demon . . . [and] if we do not know whether this evil demon exists or not, then . . . we know nothing except those things about which we are incapable of being mistaken. (McGrath, p. 55)

In this section I shall make some general remarks concerning all three objections. The next three sections will be devoted to a discussion of [1], [2], and [3] respectively.

As presented here McGrath's objections invite a facile reply. McGrath seems to make two claims: (1) Existence claims for which we have no evidence (e.g., 'there are leprechauns'; 'there are hobgoblins'; 'there exists a God'; 'there exists a Cartesian evil demon'; etc.) are counter examples to PD; for we ought to disbelieve them *even though we have no reason to believe that we are in good epistemic position relative to them.* (2) PD requires us to suspend judgment about existence claims for which we have no evidence and relative to which we have no reason to believe ourselves to be in GEP. Both claims are false. PD merely specifies a *sufficient condition* for disbelieving P. McGrath's purported counter-examples fail because they only show that being in GEP relative to P is not a necessary condition for disbelieving P. McGrath's second claim is false because from the fact that the antecedent of PD is not satisfied for a particular existence claim P *nothing* follows about which attitude we should adopt towards P. Therefore PD does not require us to suspend judgment about existence claims relative to which we are not in GEP. Exit McGrath.

A more interesting interpretation of McGrath's objections results if we make the following plausible conjecture: in making these objections McGrath was assuming that Morris accepts a *Default Assumption* (DA) governing our assessment of positive existence claims in the absence of any evidence for them:

DA Unless PD requires disbelief, suspend judgment about unsupported existence claims.

PD conjoined with DA yields the *Augmented Principle of Disbelief*, APD. For the sake of this discussion I shall henceforth assume that Morris holds APD and that McGrath attacks the adequacy of APD. APD *does* require us to *suspend* belief about existence claims for which we have no evidence and relative to which we are not in GEP. This vindicates McGrath's second claim. The purported counter examples of McGrath's first claim are now best viewed as challenging the adequacy of APD. A principle for assessing unevidenced positive existence claims is adequate only if it requires us to disbelieve certain paradigmatic existence claims that every rational person should disbelieve. McGrath argues that APD fails to require *disbelief* in the paradigmatically unfaithworthy existence claims he presents as counter examples; therefore APD is inadequate.

IV

In his first objection McGrath criticizes APD for its failure to require disbelief about a class of claims that every rational person should disbelieve: claims concerning the existence of assorted 'preternatural' beings.

I concur with McGrath that a satisfactory principle for assessing unevidenced positive existence claims ought to require disbelief in unevidenced claims concerning the 'preternatural'. However, his contention that APD fails the test is wrong. For labelling something 'preternatural' does not automatically guarantee that we have no reason to believe that we are in GEP relative to it. Take, for example, leprechauns. Like McGrath they live in Ireland and the 15th edition of the *Encyclopedia Britannica* has this to say about them:

> In Irish folklore, fairy in the form of a tiny old man often with a cocked hat and leather apron. Solitary by nature, he is said to live in remote places and to make shoes and brogues. The sound of his hammering betrays his presence. He possesses a hidden crock of gold; if captured and threatened with bodily violence, he might, if his captor keeps his eyes on him, reveal its hiding place. But usually the captor is tricked into glancing away, and the fairy vanishes.

Judging from this account of leprechauns our epistemic position relative to the claim that leprechauns exist is surely no worse than our epistemic position relative to the claim that there exists a native Irish snake. Relative to both of these existence claims I could place myself into a position in which it would seem reasonable for me to believe that I was in good epistemic position relative to these existence claims. Under these circumstances APD (and PD) will require me to disbelieve unevidenced existence claims about problematic natural or preternatural entities.

V

In his second objection McGrath maintains that APD is inadequate because it requires us to suspend judgment about all pure existence claims. He claims that the notion of being in GEP is 'unusable . . . so far as pure existence claims are concerned' (McGrath, p. 56). He argues that to be in GEP relative to a pure existence claim is to possess *all* evidence concerning it. And while 'it is theoretically possible for someone to possess all the evidence . . . one could never be in a position to assert that this is all the evidence there is' (McGrath, p. 56).

We may never have reason to believe ourselves to have *all* the evidence for a pure existence claim; but this is irrelevant. S stands in GEP relative to P just in case S would have *some* evidence for P if P were true. If S possesses some evidence for P, i.e. some reliable indicator of P's truth, then, *pace* McGrath, S no longer is 'in a poor position for assessing its truth' (McGrath, p. 56). Pure existence claims relative to which S can reasonably believe herself to stand in GEP abound; 'Tigers exist' is one example McGrath considers.

Thus McGrath fails to show that the notion of being in GEP is inapplicable to any pure existence claim.

But perhaps McGrath is making a different point. He might take the inapplicability of the notion of being in GEP with regard to pure existence claims to consist in the fact that we are not in GEP relative to a single pure existence claim for which we have no evidence. So ADP does not require us to disbelieve a single pure existence claim. Therefore ADP is inadequate. Two replies come to mind. First, there are counter examples. Consider "super quasars". They are like quasars, only so much brighter that they are observable throughout the universe. This feature of theirs places us into GEP relative to the claim that there are "super quasars". Since we have no evidence for them, ADP requires us to disbelieve the pure existence claim that there are "super quasars". Second, ADP's failure to require disbelief in a wide range of pure existence claims does not render it inadequate. The existence claims about preternatural beings that McGrath used to test the adequacy of ADP are not really pure; explicit reference to our planet is usually suppressed because of it's obviousness. A rational person should disbelieve that there are unicorns, hobgoblins, leprechauns, etc. *on earth* – just as required by ADP. But if these claims are construed as strictly pure, it is no longer clear that the epistemically responsible attitude to take towards them is that of disbelief. Suspension of judgment seems to be a more adequate response to them.

To sum up. McGrath's claim that the notion of GEP is inapplicable to *all* pure existence claims rests on his mistaken view that we have to possess *all* evidence relative to a pure existence claim in order to be in GEP relative to it. It is correct that we are not in GEP relative to a great number of pure existence claims about which ADP requires us to suspend judgment. But it does not follow that ADP is inadequate because these pure existence claims are not ones that a rational person should disbelieve.

VI

In his third objection McGrath argues that since we are not in GEP relative to the claim that the evil demon exists, ADP requires us to remain sceptical about his existence. And from this APD-induced scepticism about the evil demon it is but a short step to almost total scepticism.

But McGrath does not explain the short step that takes us from not knowing that the evil demon does not exist to knowing next to nothing. Under the assumptions that (i) an externalist theory of knowledge is correct and that (ii) knowledge is not closed under known logical implication this step can effectively be blocked. McGrath failed to show that these two assumptions are wrong and therefore failed to establish that accepting APD leads to almost total scepticism.

If 'we do not know whether this evil demon exists or not, then we do not know whether or not we are being constantly being deceived. And this is equivalent to saying that we know nothing...' (McGrath, p. 55). The externalist replies that knowing that P consists in the holding of certain external relations between my belief that P and P. Knowing that these relations obtain is not a condition on my knowing that P. Consequently, my failure to know whether or not I am being deceived does not show that I know nothing.

But there is another route connecting APD-induced scepticism about the existence of the evil demon with near universal scepticism. McGrath might have taken the following line: knowledge is closed under known logical implication: $[K_S p \ \& \ K_S(p \rightarrow q)] \rightarrow K_S q$. We know that P. We know that knowing that P entails that we are not being deceived by the evil demon. It follows from the closure principle that we know that we are not being deceived by the evil demon. But because APD forces us to suspend judgment about the existence of the evil demon, we do not know that we are not being deceived by the evil demon. Therefore we do not know that P. However, this argument is no better than the closure principle appealed to. Recent work in epistemology has shown that the principle is problematic (see, for example, Robert Nozick, *Philosophical Explanations*: Oxford 1981, pp. 204–11). This suffices to show that APD-induced scepticism about the existence of the evil demon does not, by itself, lead to almost total scepticism.

VII

Having shown that McGrath's three objections against APD fail, I conclude that (i) APD is an adequate principle for assessing unevidenced positive existence claims because it is powerful enough to require disbelief in those existence claims that a rational person should disbelieve; (ii) APD is general in the sense that it is applicable to *all* (i.e. restricted and pure) positive existence claims; (iii) APD does not inevitably lead to the absurd consequence of almost total scepticism.

Morris and McGrath discussed APD with a single question in mind: Is atheism obligatory in the absence of evidence for God's existence? Building upon his critique of APD McGrath has argued that absence of evidence for God is a sufficient condition for making atheism obligatory. We can no longer endorse his seductively simple answer. Given that we have no evidence for God, atheism is obligatory only if we have a reason to believe that we stand in GEP relative to the claim that God exists. But such a reason will be hard to come by.[1]

University of Arizona,
Tucson, AZ 85721, U.S.A.

© LEOPOLD STUBENGERG 1988

[1] Many people helped me in writing this paper. I am particularly indebted to Marian David, Keith Lehrer, Peter Smith, and Scott Sturgeon.

PUTNAM'S MIRACULOUS ARGUMENT

By GARY ISEMINGER

I

PUTNAM claims that 'the supposition that we are actually brains in a vat . . . cannot possibly be true, because it is, in a certain way, self-refuting'.[1] In preparing the ground to argue for this claim he identifies two different kinds of self-refutation.

> A 'self-refuting supposition' is one whose truth implies its own falsity. For example, consider the thesis that *all general statements are false*. This is a general statement. So if it is true, then it must be false. Hence, it is false. Sometimes a thesis is called 'self-refuting' if it is *the supposition that the thesis is entertained or enunciated* that implies is own falsity. For example, 'I do not exist' is self-refuting if thought by *me* (for any '*me*'). So one can be certain that one oneself exists, if one thinks about it (as Descartes argued). (pp. 7–8)

Now it is somewhat unclear which form of self-refutation Putnam takes the supposition that we are brains in a vat to exhibit. His initial statement of the claim, immediately following the passage just quoted, suggests that it is the second:

> *If we can consider* [my emphasis] whether it [viz. the supposition that we are brains in a vat] is true, then it is not true...Hence it is not true. (p. 8)

But a later summary of the argument suggests the first:

> In short, if we are brains in a vat, then 'We are brains in a vat' is false. So it is (necessarily) false. (p. 15)

In this note I shall argue that the argument is unsuccessful whichever way it is construed.

---

[1] Hilary Putnam, *Reason, Truth and History* (Cambridge, 1981), p. 7. Further references to this book will appear in parentheses in the text.

II

Consider first the claim that the supposition that we are brains in a vat is self-refuting in the first way. This is, in effect, the claim that, if we substitute the sentence 'We are brains in a vat' for 'P' in the (clearly valid) schema 'If P then not-P; therefore not-P', the premiss will be true and the argument sound.[2]

In order to see why this will not work it is necessary to consider on what grounds Putnam says that 'if we are brains in a vat, then "we are brains in a vat" is false'. His reasons have to do with his claim that what he calls 'magical' theories of reference are wrong, so that 'one cannot refer to certain kinds of things, e.g. *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described' (p. 16). Thus

> 'vat' refers to vats in the image in vat-English, or something related (electronic impulses or program features), but certainly not to real vats, since the use of 'vat' in vat-English has no causal connection to real vats.... It follows that if... we are really the [*sic*] brains in a vat, then what we now mean by 'we are brains in a vat' is that *we are brains in a vat in the image* or something of that kind... So, if we are brains in a vat, then the sentence 'We are brains in a vat' says something false... (pp. 14–15)

It is unnecessary for my purposes to explain or assess Putnam's causal account of reference or its application to this situation. It is sufficient to note that, supposing he is right on these matters, the reasons he gives for saying that 'if we are brains in a vat, then the sentence "We are brains in a vat" is false' *are, at the very same time, reasons for denying that this sentence can be the premiss of an argument exemplifying the form of the consequentia mirabilis.*

For Putnam's point is that if we are indeed brains in a vat, then when we utter the sentence 'We are brains in a vat' what we say is something *other than* that we are brains in a vat, something (e.g. that we are brains in a vat in the image) that can, without contradicting the claim that we are brains in a vat, be false (and which, if we *are* (real) brains in a (real) vat, *is* false). But under these circumstances to suppose that the premiss 'If we are brains in a vat, then "We are brains in a vat" is false' exemplifies the form 'If P, then not-P' is simply to commit a fallacy of equivocation. Logic students learn early on that they cannot simply mechanically replace type-identical stretches of language, whether sentences, predicates, or names, with identical stand-ins in finding forms of arguments, and this is precisely a case where, the sentence 'We are brains in a vat' having different meanings in its two occurrences in the premiss, doing so is illegitimate. On Putnam's assumptions the supposition that we are brains in a vat cannot be shown to be self-refuting in the first way.

---

[2] This way of regarding the argument involves taking it as an instance of (at least a close relative of) the *consequentia mirabilis* of the medievals, known to the ancients as well. See William and Martha Kneale, *The Development of Logic* (Oxford, 1962), pp. 172–4. Interestingly, Bertrand Russell also offers an argument instantiating this form which, like Putnam's, purports to refute something called 'realism', thus: 'Naive realism, if true, is false; therefore it is false' (*An Inquiry into Meaning and Truth* (London, 1962), p. 13). There is obviously a powerful temptation to try to show something interesting in such a neat and surprising way, but in my view Russell does not succeed either. See Gary Iseminger, 'Russell's Much-Admired Argument Against Naive Realism', *Journal of Indian Council of Philosophical Research,* IV, 1 (1986), pp. 173–6. Even more interestingly, Plantinga suggests (though he does not develop the suggestion) that Putnam's own anti-realist views are vulnerable to such a refutation. See Alvin Plantinga, 'How To Be an Anti-Realist', *Proceedings and Addresses of the American Philosophical Association,* 56, 1 (1982), p. 64.

III

Consider now the possibility that the supposition that we are brains in a vat is self-refuting in the second way.[3] If the argument is construed this way, its form is something like 'If we can think that P, then not-P, and we can think that P; therefore not-P'.

Now this is plainly a validating form. Furthermore, Putnam's argument for the first premiss of the instance in question (the premiss that if we can think that we are brains in a vat, then we aren't) depends on some of the same considerations just adduced in the discussion of the possibility that the supposition is self-refuting in the first way. That is, according to Putnam, because of the absence of appropriate causal connections between their thoughts and vats, brains in vats cannot refer to vats and cannot, therefore, have the thought that they are brains in vats. So let us grant Putnam the premiss that if we can think that we are brains in a vat, then we aren't.

Now note, however, that, on the current reading, unlike the reading of the argument as exemplifying the *consequentia mirabilis*, *there is another premiss to the argument*, the premiss that we can think that we are brains in a vat. Is Putnam entitled to this premiss?

The comparison with Descartes (with 'P' taken as 'I don't exist' instead of 'We are brains in a vat' and the 'we can think's changed to 'I am thinking's in the schema given at the beginning of this section) might suggest that there is no special problem here. One of the apparent strengths of the argument Putnam seems to attribute to Descartes is that the analogous premiss in that argument (that he is entertaining the thought that he does not exist) seems to be knowable by some kind of introspective reflection. Is it not similarly introspectively evident to us that we can entertain the thought that we are brains in a vat? What else have we been doing for the last few minutes?

---

[3] If one has to choose, there is perhaps more textual evidence for taking this to be Putnam's claim than there is for taking his claim to be that the supposition is self-refuting in the first way. For example, when he first introduces the argument he poses the question to which he wants to give a negative answer thus: 'Could we, if we were brains in a vat . . . , *say* or *think* that we were?' (p. 7).

But here, as before, the very considerations which support the conditional premiss of the argument undermines the rest of it, though in a different way. The problem in this case is not, as it was before, the *logical* one that we can no longer make the argument exemplify the relevant form (nor is it, so far as anything I will argue here is concerned, the *alethic* one that not all the premisses are true). The problem now is rather the *epistemic* one that we cannot without begging the question help ourselves to the second premiss if we accept Putnam's argument for the first. For given the causal account of reference which Putnam invokes to defend the first premiss, we *cannot* in general introspectively discern what thoughts we are having; (not only God but) *we* cannot look into our minds to see there what we are speaking of. Given that, so far as anything 'in the head' is concerned, we *are* in an appropriate state to be thinking that we are brains in a vat, what more would be required to assure us that we actually *are* entertaining that thought? Evidently, on Putnam's view, the knowledge that we stand in 'appropriate' causal relations to vats. But, if Putnam is right in his defence of the *first* premiss, we can only have *that* knowledge if we know that we are *not* brains in a vat, and *this* knowledge, of course, is what the argument is supposed to *give* us, not something we can invoke in defending one of its premisses.

It is important to see that this is not merely a variation on the old saw about the syllogism begging the question. I am *not* claiming that we *never* have an epistemic right to assert the premisses of a valid argument prior to our having an epistemic right to assert its conclusion. My objection would not, for example, tell against Putnam's version of the Cartesian argument, where our epistemic right to the first premiss would presumably be underwritten by some thesis about the necessary connection between entertaining thoughts and existing and our epistemic right to the second would be underwritten by introspection. My point against Putnam is that, given his views about the conditions for reference and about the relationship between what we are referring to and the contents of our thought (views which, as it happens, are his reasons for asserting the first premiss of *his* argument), we do *not* have the same epistemic right to assert that we are thinking that we are brains in a vat that (at least on Descartes's view) we have to assert that we are thinking that we do not exist.[4] Further, on these views about reference and the contents of thoughts, precisely what we need to know in order to assure ourselves that we *are* entertaining the thought that we are brains in a vat is what the argument is trying to *show* us, that we are *not* brains in a vat.

---

[4] This may be true not only on Descartes's 'magical' view of reference but even on Putnam's causal view, for it seems plausible to suggest that brains in vats stand to themselves (and can even know introspectively that they stand to themselves) in whatever causal relations are necessary to enable them, even on Putnam's view, to refer to themselves and thus to have the thought that they exist and know that they are having it. At least *that* case of meaning *might* be 'in the head'.

Putnam's summary of his version of Descartes's argument is that 'one can be certain that one oneself exists, if one thinks about it' (p. 8). Perhaps a better way to put it would be that one can be certain that one oneself exists, if one *can be certain that* one is thinking about it. Similarly, we can be certain that we are not brains in a vat, if we can be certain that we are thinking about it. But, it seems plausible to say, we *can*, without begging the question as to whether or not we exist, be certain that we are thinking about our existence. On the other hand, given Putnam's views about reference and content, we *cannot*, without begging the question as to whether or not we are brains in a vat, be certain that we are thinking about our being brains in a vat. But then the claim to parallelism with the Cartesian argument as Putnam interprets it, and with it Putnam's argument construed as an attempt to show that the claim that we are brains in a vat is self-refuting in the second way, fails.

I conclude that Putnam's attempt to show that the claim that we are brains in a vat is self-refuting is self-defeating.

IV

I have not said anything about the larger context in which Putnam embeds his views on the possibility of our being brains in a vat, his attempted refutation of the realist view that he calls 'Externalism'. It is not at all clear just how the claim that we couldn't be brains in a vat is supposed to figure in this refutation.[5] Despite Putnam's repeated claims to have shown that we couldn't be brains in a vat, however, it seems to me at least possible that an argument can be extracted from the text (especially pp. 49–51) which moves from the denial of the 'magical' theory of reference *via* reflections on the situation of brains in vats to the denial of Externalism *without* moving through the intermediate conclusion that we couldn't be brains in a vat. This argument might require as an intermediate step only the (much weaker) claim that *if* we are we can't think that we are. Whether such an argument is discernible in the text and, if so, whether it is ultimately successful are stories for another occasion, but such an argument would, of course, not be vulnerable to the objections I have brought against Putnam's miraculous argument.[6]

---

[5] One notable expression of puzzlement on this score is in David Lewis, 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62 (1984), pp. 233–6.

Carleton College,
Northfield, MN 55057, U.S.A.

© GARY ISEMINGER 1988

---

[6] I am grateful to my colleagues P. C. Mason and Fred Stoutland for their helpful comments. The Editor has drawn to my attention an article I had not seen, Jane McIntyre's 'Putnam's Brains', *Analysis* 44.2, March 1984, pp. 59-61, which criticizes Putnam on grounds similar to mine.

REALISM WITHOUT TRUTH

By ROD BERTOLET

ONE of the most distinctive features of Michael Devitt's defence of scientific realism is his claim that it can be carried out without claiming that current scientific theories are true, or approximately true.[1] I shall argue that this feature of his approach undermines his defence of scientific realism.

Realism, for Devitt, is a doctrine about the nature of things — or rather two doctrines, one about the nature of the observable things posited by common sense and another about the nature of the unobservable things posited by scientific theories. The two views are characterized thus:

> *Common-Sense Realism* Tokens of most current observable common-sense, and scientific, physical types objectively exist independently of the mental.
> *Scientific Realism* Tokens of most current unobservable scientific physical types objectively exist independently of the mental. (p. 22)

These characterizations bring out the two dimensions of realism on which Devitt insists: that certain sorts of entities exist objectively, and that they exist independently of the mental (see sections 2.2 and 2.3). Let us turn to the arguments for them.

Both arguments are inferences to the best explanation. The premisses of the argument for Common-Sense Realism are various *as-if* statements — 'It is as if there is a black raven on the lawn', etc. — and the conclusions are various *that's the way it is* statements — 'There is a black raven on the lawn', etc. More carefully, statements of the first sort ('It is as if . . .') are generally best explained by statements of the second sort ('There really is/are . . .'), so, by inferences to the best explanation, we are entitled to believe that there really are trees, tables, chairs, and so on, where these objectively exist independently of the mental. This is not to rule out occasional aberrations; it may be that the best explanation of why it is as if there are pink rats cavorting before *S* is that *S* has ingested an unwise quantity of alcohol. Let us suppose that this is a successful defence of Common-Sense Realism. (While I have some reservations about Devitt's response to scepticism, I think it *is* or can be worked into a successful defence. But my argument does not turn on whether this is correct; my point is that even if this defence of Common-Sense Realism succeeds, the defence of Scientific Realism does not.)

---

[1] Michael Devitt, *Realism and Truth*, Oxford: Basil Blackwell, 1984. All references are to this book.

The argument for Scientific Realism is also an inference to the best explanation. Instead of being an inference (or set of inferences) from 'as-if' talk to 'really is' talk, this one is from observables talk to unobservables talk. More carefully, the argument is that if we assume that the confident and necessary posits of our favoured theories exist, we can give good predictions and explanations of the behaviour of observed entities; indeed without assuming this we have *no* explanation of such behaviour (and this is basically the argument against the classical alternatives to Devitt's position — see chapter 7). Since this is the best explanation of the behaviour of observable things, we should believe that such unobservables exist.

Unlike (say) Putnam and Boyd, who want to argue for realism on the basis of our having gotten better and better theories about the world — the doctrine of convergence — Devitt does not want to rely on any premisses concerning the truth or even approximate truth of current theory. As he notes earlier, in discussing what he takes to be constitutive of realism, the ontological intuitions of realism

> do not require us to be committed even to the approximate truth of statements other than existence statements, and . . . do require us to be committed to the *full* truth of *most* of the existence statements. (p. 33)

Not only do the ontological intuitions of realism not require this, but 'It is foolhardy to be committed to the truth of even the best scientific theories, as the history of science shows' (pp. 37–38). The idea is thus that while we must accept the ontology of our theories (or rather the necessary and confident posits, leaving behind those about which we are not so sure), we neither need nor ought to want to accept the theories themselves, i.e. be committed to the truth of them. The same point is made earlier, when Devitt quite explicitly rejects the claim that we should say that tokens of scientific types must have some of the properties they are believed to have, adopting instead a Kripke-inspired form of *a posteriori* essentialism on which 'It would make realism unnecessarily strong to commit it not only to the existence of entities of certain kinds but also to our being right about their properties' (p. 19). What I want to suggest is that this deprives us of the claim that allegiance to the unobservables of scientific theory is the best (only) explanation of the behaviour of observables, that we *must* pledge allegiance to more of those theories than such existence statements. And I shall also suggest that some of Devitt's own remarks support this claim.

The trouble is that *that there are electrons* (etc.), by itself, doesn't explain anything at all about the behaviour of observable objects. What explains the behaviour of the observable computer before me is that there are electrons, protons, neutrons, etc. *that obey such-and-such laws.* These laws are just those postulated by our best current theories. If we are to get any explanations of the behaviour of the computer and other observables, we must appeal to such laws – hence, accept at least large parts of current theory *beyond* existence statements as true. (I can see but one way around this claim: we could argue that it is part of the meaning of 'electron' that it obeys certain laws. Devitt would obviously, and I think quite rightly, reject this move.) So I do not see how Devitt can avoid including a statement of the laws that we believe unobservables obey in the premisses of the inference to the best explanation, which is to say, I do not see how he can avoid commitment to the truth of our most confidently held theories.

Oddly – or perhaps reasonably – enough, Devitt says as much in discussing theoretical success.

> If there are electrons, *and they are the way electron theory says they are,* and such and such experimental conditions obtain, then so and so observational results should occur. If it does occur, that is a success of electron theory ... (p. 109, emphasis added).

Just so. But this just reinforces the point that it cannot be that the behaviour of observables is explained *solely* by existence statements. Appeal must be made to the truth or approximate truth of our theories. And if it is approximate truth that is required, then more work on verisimilitude may also need to be done in order for the argument to be successful. (The truth or verisimilitude on which one may need to rely might be disquotational rather than explanatory, so it may turn out to be possible to respect Devitt's wish that we 'not take truth for granted' even if we must say that current science is true, or approximately true – see Maxim 4, p. 4, and subsequent discussion.)

I do not mean to suggest that we ought to embrace as true all available scientific theories; Devitt is surely right to say that going that far would be foolhardy in light of the history of science. All that the argument requires is acceptance of what Devitt might call the confident and necessary laws of our confident and necessary theories, not allegiance to everything tried out in the physics journals; and this of course is compatible with a fallibilist attitude towards even these 'confident and necessary' laws.

Do the considerations that force us to appeal to the truth of more of our scientific theories than their existence statements force a change in Devitt's argument for Common-Sense Realism? I think not. It is not that we infer from as-if statements such as 'It is as if there are cats' or 'It is as if there are trees', that 'There are cats' or 'There are trees'. Rather, the premisses are statements such as 'It is as if there is a black raven on the lawn', from which we conclude 'There is a black raven on the lawn'. It is then not the bare existence of ravens, etc. that is involved; rather it is claimed that there being certain objects *with certain properties* is the best explanation of its being as if there are certain objects with certain properties. But even if this is acceptable for the argument for Common-Sense Realism, Devitt cannot make a similar move on behalf of Scientific Realism. In the argument for Common-Sense Realism what is inferred is again the existence of objects *with certain properties*. An analogous argument to the best explanation in defence of Scientific Realism would require Devitt to suppose that it is the existence of e.g. electrons *with certain properties* that explains the behaviour of observables. But to suppose that it is the existence of e.g. electrons *with certain properties* that explains the behaviour of observables is, if these properties are to have any explanatory value, to suppose that electron theory is substantially *true* — which is precisely the supposition Devitt wishes to avoid. The only conclusion I can see is that the most distinctive feature of Devitt's realism, the attempt to argue for Scientific Realism without claiming that at least part of current science is correct or nearly correct, precludes the argument from being successful.

Purdue University,
West Lafayette, IN 47907, U.S.A.

© ROD BERTOLET 1988

REALISM WITHOUT TRUTH: A RESPONSE TO BERTOLET

By MICHAEL DEVITT

ROD Bertolet's helpful comments [1] reveal an error in my discussion of scientific realism in *Realism and Truth* [2]. However, I think that he misdescribes the error and is likely to mislead people as to its significance for a central thesis of the book: the irrelevance of correspondence truth to realism.

I emphasize two distinctions that are important to the assessment of Bertolet's criticism. The first distinction is between two notions of truth. *Disquotational* truth is a non-explanatory logical 'device for talking about non-linguistic reality while appearing to talk about sentences' ([2], p. 29; see also [4], pp. 162–5). *Correspondence* truth (and *vertificationist* truth, for that matter) is an explanatory notion in a substantive theory of language or mind. The second distinction is between the *constitutive* issue of what realism is and the *evidential* issue of why we should believe in it (Maxim 1, [2], p. 3).

Bertolet says that 'one of the most distinctive features of Michael Devitt's defence of scientific realism is his claim that it can be carried out without claiming that current scientific theories are true, or approximately true' ([1], p. 195). THe view attributed to me clearly concerns the evidential issue. However, the three passages Bertolet quotes in support ([2], pp. 38, 37–8, 19) do not justify the attribution because they are all drawn from discussions of the constitutive issue. Furthermore, only if the attribution concerns correspondence truth not disquotational truth is it justified by any other passages in the book. I shall discuss the evidential issue later.

I shall start by discussing the constitutive issue. I am concerned that Bertolet's use of passages on that issue to support an attribution on the evidential issue may lead to the confusion of my claims about the bearing of truth on the two issues. It seems to me that my view of the bearing of truth on the constitutive issue is much the more distinctive of the two.

I THE CONSTITUTIVE ISSUE AND CORRESPONDENCE TRUTH

I argue that realism should be a metaphysical (ontological) doctrine and that correspondence truth is not in any way constitutive of it: you can and should say what realism is without appeal to any explanatory semantic notion (Maxim 2, [2], p. 3). Accounts of realism are usually dominated by talk of truth, reference and convergence to the point where it can be hard to notice the metaphysical issue. (See, for example, Jarrett Leplin's editorial introduction to [6].) I argue further that we should settle the metaphysical issue before any semantic one (Maxim 3, [2], p. 4). The dominance of semantics over metaphysics has been very detrimental to realism. (Note, for example, that Laudan's [5] is often thought to be very damaging to realism even though it is almost entirely about truth, reference and convergence.)

Here is one way of bringing out the irrelevance of correspondence truth to the metaphysical issue. Consider this fairly typical example of a contemporary semantic version of scientific realism:

> *Contemporary Realism* Most scientific statements about unobservables are approximately correspondence-true.

Why would people believe *Contemporary Realism*? I suggest only because they believed something like the following two doctrines:

> *Strong Scientific Realism* Tokens of most unobservable scientific types objectively exist and approximately obey the laws of science;
>
> *Correspondence Truth* Sentences have correspondence-truth conditions (presumably explained in terms of objective reference relations).

The 'equivalence thesis' should hold for correspondence truth as for any other truth. So appropriate instances of

> *s* is approximately correspondence-true if and only if approximately *p*

will hold. *Strong Scientific Realism* and *Correspondence Truth*, together with these instances, imply *Contemporary Realism*. I can see no reason for believing *Contemporary Realism* that does not involve, at least implicitly, *Strong Scientific Realism* and *Correspondence Truth*.

Strong Scientific Realism is a metaphysical doctrine about the underlying nature of the world in general, a largely impersonal non-linguistic world. To accept this doctrine we have to be confident that science is discovering things about the unobservable world. Does the success of science show that we can be confident about this? Is inference to the best explanation appropriate here? Should we take sceptical worries seriously? These are just the sort of epistemological questions that have been, and still largely are, at the centre of the realism debate. *Their home is with Strong Scientific Realism not with Correspondence Truth.*

Correspondence Truth is a semantic doctrine about the pretensions of one small part of the world to represent the rest. The doctrine is the subject of lively debate in the philosophy of language, the philosophy of mind and cognitive science. Do we need that notion of truth in a theory of linguistic symbols and mental representations? Should we prefer a conceptual-role semantics? Or should we, perhaps, near enough eliminate meaning altogether? These are interesting and difficult questions (Maxim 4, [2], p. 4) but they bear no special relevance to the metaphysical concerns of *Strong Scientific Realism* ([3], sects. 2–3). They are not even particularly concerned with the theoretical language of science.

Strong Scientific Realism and *Correspondence Truth* have very different subject matters and should be supported by very different evidence. Underlying *Contemporary Realism* is a conflation of these two doctrines that has been detrimental to both. The conflation has brought irrelevant epistemological arguments against *Correspondence Truth* ([2], pp. 150–2, 157–61, 193–5; [3], sects. 5–7). It has made adherence to *Strong Scientific Realism* seem comical in some circles when it should seem natural and appropriate.

II THE CONSTITUTIVE ISSUE AND DISQUOTATIONAL TRUTH

Though correspondence truth should have no role in stating realism, disquotational truth may have, for disquotational truth may have a role in stating any theory. This role can be completely trivial: instead of asserting 'T', assert ' "T" is true'. Even where the role is not trivial, it does not make the theory in any way semantic. (I agree with Putnam that Tarski's famous article would have been better called 'The *Non*semantic Conception of Truth'; [7], p. 63.)

In the first passage Bertolet quotes to support his attribution (from [2], p. 38), my concern is, explicitly, with disquotational truth and with doctrines that I have earlier defined and named, *Common-Sense Realism* and *Scientific Realism* ([2], p. 22). The latter, amended for our purposes, is as follows:

> *Scientific Realism* Tokens of most unobservable scientific types objectively exist.

I point out that *Scientific Realism* requires the disquotational truth of most of the existence statements of science but not even the approximate disquotational truth of the statements of science in general. One could, of course, define a stronger doctrine, but it would be foolhardy to commit oneself to the total disquotational truth of scientific theories. That is my point in the second quoted passage ([2], pp. 37–8). However, there are intermediate positions between the foolhardy view and my *Scientific Realism*. *Strong Scientific Realism* above is an example, as can be seen if we rewrite it using disquotational truth:

> *Strong Scientific Realism** Most scientific statements about unobservables are objectively approximately disquotational-true.

My only reason for not urging *Strong Scientific Realism** was a tactical one. Such a doctrine, being stronger, is harder to defend than *Scientific Realism* ([2], p. 19). I thought that my weaker doctrine was strong enough to counter the many contemporary anti-realisms (van Fraassen, Kuhn, Feyerabend, Putnam, Goodman, Dummett, etc.). That is what underlies the third quoted passage ([2], p. 19). I certainly do *not* think that 'we neither need nor ought to want to . . . be committed to the truth' of scientific theories ([1], p. 196).

III THE EVIDENTIAL ISSUE AND DISQUOTATIONAL TRUTH

My error on the evidential issue was not that I claimed that we could defend realism without appeal to the disquotational truth of scientific theories; indeed, in effect, I do appeal to this in my defence of *Scientific Realism* ([2], pp. 108–9), as Bertolet notes. I argue that the main reason for believing in the existence of unobservables is that by so doing 'we can give good explanations of the behaviour and characteristics of observed entities' ([2], p. 104). But, as Bertolet emphasizes ([1], p. 197), the goodness of these explanations depends on the unobservables obeying the laws of science. My error was in not making this much more explicit and in not noting that it diminishes the above-mentioned tactical advantage of defending the weaker doctrine. I say 'diminishes' not 'destroys' because I think it arguable that when in the past we have been fairly wrong about the nature of entities, hence fairly *wrong in our reasons* for positing them, we have mostly been *right* to posit them ([2], pp. 143–9). So rather less disquotational truth may be required for defending *Scientific Realism* than is entailed by *Strong Scientific Realism**.

IV The Evidential Issue and Correspondence Truth

So I agree with Bertolet that the defence of *Scientific Realism* requires a deal of truth in scientific theories. However, where he allows parenthetically that the notion we need *may be* only disquotational truth ([1], p. 197), I want to insist that the one we *need* is only disquotational truth. Nothing in the argument for unobservables that Bertolet rightly emphasizes requires correspondence truth, or indeed any other notion from a semantic theory. Correspondence truth is not only not constitutive of Realism, it is not essential to its defence.

This is not to say that there are no arguments for *Scientific Realism* that involve correspondence truth and related notions of convergence and verisimilitude. I think that there are ([2], pp. 38–40, 114–17, 156–7, 229–31), but these are not essential. The essential argument is the one Bertolet and I give. To argue from semantics to ontology is to put the cart before the horse.

In conclusion, correspondence truth has no role in stating scientific realism and little in its defence. Disquotational truth can have a role in stating realism, but this is of no significance. I erred in underplaying the importance of the disquotational truth of theories to the defence of realism.[1]

University of Maryland,
College Park, MD 20742-7615, U.S.A.

© Michael Devitt 1988

References

[1] Rod Bertolet, 'Realism without Truth', *Analysis* 48.4 (October 1988), 195–8.
[2] Michael Devitt, *Realism and Truth*, Oxford: Basil Blackwell, 1984.

[1] My thanks to Fiona Cowie for comments on a draft.

[3] Michael Devitt, 'Rorty's Mirrorless World', *Midwest Studies in Philosophy, Volume XII: Realism*, Peter French, Howard Wettstein, and Theodore Uehling, eds., Minneapolis: University of Minnesota Press, 1988, 155–77.
[4] Michael Devitt and Kim Sterelny, *Language and Reality: An Introduction to the Philosophy of Language*, Oxford: Basil Blackwell, 1987.
[5] Larry Laudan, 'A Confutation of Convergent Realism', *Philosophy of Science* 48 (1981), 19–49. Reprinted in [6].
[6] Jarrett Leplin, ed. *Scientific Realism*, Berkeley: University of California Press, 1984.
[7] Hilary Putnam, 'A Comparison of Something with Something Else', *New Literary Theory* XVII (1985–6), 61–79.

CLOSED TIME AND CAUSAL LOOPS: A DEFENCE AGAINST MELLOR

By SUSAN WEIR

I

AT seven o'clock one sunny evening, a man applies a match to the charcoal in his barbecue. A second or two later, the charcoal ignites. This is, in principle, a highly probable causal chain. At the same moment, a neighbour applies a match to her barbecue in order to make it light at the earlier time of 6 p.m. This is not a possible causal process, as plausibly argued by D. H. Mellor [1]. Finally, in a third well-kept garden, a match is applied to the barbecue at seven o'clock and, as a result, it ignites at an earlier time, not just a few seconds, but millions of years before. This, I shall argue, is perfectly possible and, within the appropriate context, is left quite untouched by Mellor's arguments against the second case.

Mellor's intention in Chapter X of *Real Time* is to disprove the possibility of backward causation. In order to do this he argues that causal loops cannot fulfil at every stage the requirements for being a causal process. His example is of a small, localized loop like that in Case Two above. He claims, though, that his arguments apply to all causal loops including ones which consist of the entire history of the universe. 'Cyclical universes are as impossible as backward causation, for the same reason' ([1], p. 187). I shall argue that, in the context of what he calls a cyclical universe, Mellor's arguments against causal loops become either irrelevant or invalid. It is in such a world that the third case described above is perfectly possible.

I shall start by describing the sort of universe in which Case Three takes place. It is in essence one which was constructed by Newton-Smith ([2], pp. 65–7). In general appearance, it is very like our own, with similar events occurring on a similar time-scale. The accepted scientific hypothesis about the history and future of the universe is the now familiar story in which the Big Bang, expansion of the universe and increase in entropy is followed by shrinkage accompanied by a decrease in entropy and, eventually, the collapse of the entire universe into one tiny volume of huge mass. This, the physicists predict, is to be followed by a Big Bang, expansion, shrinkage, total collapse and so on. Such a cyclical sequence of events is thought to extend to infinity in both past and future. A further conjecture about the nature of this universe must be added to make our picture complete. This is that not only cosmic, but all events down to the tiniest detail, run through an exactly similar sequence from Big Bang to total collapse on every cycle.

The next move in the construction of the universe of Case Three is to apply the Principle of the Identity of Indiscernibles to the events in the world described above. Since each event in a particular cycle is exactly similar to the event at the equivalent point on every other cycle and, furthermore, has exactly similar relational properties, we can conclude that there is actually only one token of each event type and, consequently, just one cycle. An event is not preceded and succeeded, at the elapse of one cycle, by another exactly similar event, but by itself. We conclude, then, that all events in this universe form one vast causal and temporal loop. It is a closed time world, the temporal relations between its events being topologically equivalent to the spatial relations between points on a circle. On a local time scale, causal and temporal relations are just as they appear to be in the real world but, on a global scale, every event both precedes and succeeds every other. As Figure 1 illustrates, if the match is applied to the barbecue at *t*, it will light a few seconds later at *t*′. However, *t*′ also precedes *t* by a duration a little less than the whole history of the universe, let us call it one billion years less five seconds. Therefore the application of the match at *t* also causes the charcoal to ignite one billion years minus five seconds before *t*. It may be argued that such a structure for time is impossible or even illogical. In this paper I aim to show only that the causal loop that is entailed by it is not inconsistent in the ways suggested by Mellor. I have, however, replied to the many objections to a closed time universe in Weir [3].

Before considering Mellor's arguments, it will be useful to distinguish two senses of backward causation. In one sense, it is a violation of the principle that a cause may not succeed its effect. In any causal loop, this principle is broken but in closed time the cause, though succeeding its effect, also precedes it. In a second sense, it is causation which goes against the general order of causation. This sense applies to a local causal loop such as in Case Two are alight are told to apply matches and half told not to apply matches, the same proportions applying to fires that have not ignited.

At the first stage of Mellor's discussion, two possible outcomes are considered. Either the igniters obey their instructions and so no correlation arises or they fail to follow instructions consistently and a positive and significant correlation between matches being applied at t_2 and fires lighting at t_0 is observed. In the first case, clearly nothing has been done to prove backward causation. In the second, according to Mellor, we still have no support for the hypothesis of backward causation because there is an explanation in terms of forward causation. We may find accidental reasons which coincidentally have resulted in the correlation (like some igniters forgetting what their instructions were, for example). Such coincidences, of course, prove nothing and must be eliminated. Once this has been done, if the correlation persists, then we must conclude, says Mellor, that the successful lighting of fires at t_0 makes more likely the application of matches at t_2. This is the only way that we can explain the failure of the igniters to carry out their instructions.

If the application of matches at t_2 were really free from the causal influence of the events at t_0, then, Mellor argues, igniters could apply matches to those fires that had not started at t_0 or refrain from igniting those that had. Once the preceding effect has been established, the succeeding cause can always be interfered with as long as it is not causally fixed beforehand. If the cause were fixed beforehand, the experiment is open to the charge that the causal process actually worked from earlier to later and thus involved only forward causation.

In the context of a closed time world, the idea of testing for backward causation as Mellor presents it, does not arise. There is no question of experimentally ascertaining that backward causation takes place in one particular case. Because of the temporal structure of such a world, all forward causation is also backward causation. Any example of causation works backwards in the sense that the cause succeeds the effect, but it also precedes the effect. As I commented earlier, the sort of backward causation entailed by closed time does not involve causal *reversal*, i.e. proceeding against the general direction of causality. If we have the evidence necessary to adopt a closed time theory, then we have the evidence for backward causation. For the same reason, Mellor's argument that the test must fail does not stand up to examination. He says that if the required correlation does arise, we must accept the alternative of an explanation in terms of forward causation. This is justified, he later claims, because 'the causal criterion for the direction of time entitles me to take forward causation for granted' ([1], p. 184). In the closed time context, however, the causal criterion for the direction of time gives us only one direction for both causal and temporal order and therefore there is no question of any alternative explanation.

III

Although Mellor's test situation does not apply in closed time, a closer look at it will further reveal just how backward causation in closed time is different from a local causal loop. The second stage of his argument is also derived from a situation where the intended cause is interfered with once the effect has been established. In this situation, when it is seen that a fire has ignited at t_0, the igniter is ordered not to put a match to it at t_2. The igniter complies with the order, thus avoiding Mellor's stage one argument. Then some change of evidence about the past arises so that it is suddenly found that in fact the fire did not light at t_0 at all. This change of evidence happens on a regular basis and thereby the correlation needed to prove backward causation is produced. Mellor argues that if this were to happen, we would have to ask ourselves why the original evidence was so unreliable. On investigation, either the faults would be found and corrected and therefore the correlation would break down again or we would find the errors impossible to account for which would lead Mellor to the third stage of his argument.

In the closed time world, such changes of evidence cannot arise since it is impossible to carry out the test of backward causation in this way. Mellor's test relies on the ability to ascertain, before instructing the igniters, whether the fires have ignited or not. The igniters having carried out their instructions, we then reassess whether the fires have lit. In the closed time world, there will be no evidence available before the application of the match as to whether a fire has lit one billion years less five seconds before. Since the time when the fire lit (or not), and before the application of the match, the entire universe has collapsed into a single mass. Nothing can preserve its identity through such a state, and so there can be no traces of events before it available to observers afterwards. Whatever reason the inhabitants of such a world might have for hypothesizing that the cycle of their universe is unvarying, it cannot be justified on the basis of evidence about what has taken place before the Big Bang, because no such evidence can exist. Evidentially based knowledge is restricted to past time after the Big Bang, and the future up to the total collapse can only be predicted. Since evidence about the past occurrence of the effect is not available, we clearly cannot test the backward causation hypothesis in the way envisaged by Mellor, but this does not adversely affect my case for backward causation in closed time because, as I argued in Section II, it is, in this context, justified in a quite different way.

IV

At the third stage of his argument, Mellor contends that regular changes of evidence would contravene his assumptions that all real events have contiguous effects and that, in principle, the most reliable perception of an event is through these effects. According to these principles, the most reliable evidence that a particular effect has occurred is available immediately after it happens and so before the occurrence of the cause. Furthermore, he says, since it is an event's effects, not its causes that affect the reliability of its perception, we cannot accept a situation in which the occurrence of the cause increases the reliability of the observation of the effect. We must be able reliably to observe the effect before the cause happens and if this is the case, the stage two argument applies. Mellor claims that:

> The only way the experiment can fail to disprove backward causation is if a fire's igniting or not at t_0 cannot be perceived at all, reliably or unreliably, until after it is seen at t_2 whether a match has been put to it . . . this means the fire's igniting or otherwise must have no effects before t_2 ([1], p. 183).

That, in turn, would mean that an event would have no contiguous effects, something which, as we have seen, Mellor assumes cannot be the case.

In the context of closed time, the condition of Mellor's hypothetical is fulfilled. Because t closely precedes t', as in Figure 1, no observer can perceive the fire's igniting until after the application of the match is perceived. Although the fire has lit (or not lit) in the past, the most reliable knowledge of this fact cannot be available now because of the vastness of the intervening duration. As Mellor says, the most reliable evidence that an event has happened is available immediately after it occurs. In addition, as we saw in Section III, although the fire had effects in the past, any traces of it have been erased by the total collapse of the universe. The most that an observer might have immediately before the instigation of the cause is a prediction of whether the fire ignites or not. Easily the most reliable knowledge of the outcome is available immediately after t' (perceived through the fire's contiguous effects) and, therefore, locally after t. Thus, in principle, we have knowledge of the effect only after we know that the cause has taken place but this does not imply the consequent of Mellor's hypothetical, that the fire's igniting can have no effects before its cause succeeds it.

Backward causation in closed time, then, *can* fulfil the conditions for being a causal process that Mellor imposes. The reason is that our most reliable knowledge about the occurrence of the effect is available only after the cause takes place. Two factors contribute to this fundamental difference between closed time and the local causal loop. One is the fact that the closed time loop takes in all the events of universal history and so the time locally after the effect is not locally before the cause and the time locally after the cause *is* locally before the effect. The other is that the total collapse of the universe which intervenes between the effect and the cause destroys all information about the effect for those observing before the cause occurs. These two points can be added to the fundamental difference discussed earlier: that, unlike in a local causal loop, backward causation in closed time is also forward causation.

Lynfield, Frome Park Road,
Stroud, GL5 3LF

© SUSAN WEIR 1988

REFERENCES

[1] D. H. Mellor, *Real Time*, Cambridge: Cambridge University Press, 1981.
[2] W. H. Newton-Smith, *The Structure of Time*, London: Routledge and Kegan Paul, 1980.
[3] S. I. Weir, *An Enquiry into the Possibility and Implications of a Closed Temporal Topology*, unpublished thesis, University of Bristol, 1986.

PERMISSIONS AND ILLOCUTIONARY ACT TAXONOMY

By CHRISTOPHER NEW

I

J. R. SEARLE remarks that the most important weakness of Austin's taxonomy of illocutionary acts in *How to do things with Words* ([1], pp. 147 ff.) is that 'there is no consistent principle of classification'; and he claims he has removed this weakness in his own taxonomy, which takes 'illocutionary point and its corollaries, direction of fit and expressed sincerity conditions, as the basis for constructing a classification' ([2], p. 12).

One of his categories is *directives*: 'the illocutionary point of these consists in the fact that they are attempts . . . by the speaker to get the hearer to do something' ([2], p. 13). It is puzzling to discover that Searle goes on to list permissions under directives whilst at the same time conceding that giving permission lacks the illocutionary point which defines that category. His attempt to surmount this apparently insuperable difficulty is revealing. 'Permit', he claims (along with 'dare', 'defy', 'challenge', 'invite' and — in one of its senses — 'advise'),

> has the syntax of directives, though giving permission is not strictly speaking trying to get someone to do something, rather it consists in removing antecedently existing restrictions on his doing it, and is therefore the illocutionary negation of a directive with a negative propositional content, its logical form is ~!(~p). ([2], p. 22)

Now these claims, whether or not correct in themselves, import the very inconsistency of classification into Searle's taxonomy which he deprecated in Austin's. Instead of his professed single criterion of illocutionary point, with its corollaries, he now suggests three criteria, and three which unsurprisingly yield quite different classifications. By the *illocutionary point* criterion neither permissions nor defyings would be directives; for permissions are not trying to get someone to do something, and defying you to do your worst is not trying to get you (or for that matter trying to get you not) to do your worst. By the *syntactic criterion*, however, which is that first person present tense sentences have the deep structure

I verb you + you Fut. Vol. Verb (NP) (Adv),

([2], p. 22) *all* the acts designated by the verbs listed above would be directives. And by the *illocutionary negation* criterion permissions would be directives, but the others would not — inviting or advising you to pay your taxes is not removing antecedently existing restrictions on your doing so. With criteria as conflicting as these we are not very likely to improve on Austin's original tentative classification.

Savas L. Tsohatzidis, in his diagnosis of Searle's problem [4], also points out that (by the illocutionary point criterion) permissions are not directives; however, he imputes a deeper confusion to Searle than that of conflicting classificatory criteria: namely, the confusion of supposing that the illocutionary negation of a directive is, like the directive negated, an attempt to get the hearer to do something. On this view we must take Searle to believe that to utter

(1) I do not order you not to ϕ

or

(2) I do not forbid you to ϕ

each of which he considers equivalent to

(3) I permit you to ϕ

([2], p. 22; [3], p. 5, p. 202), is to attempt to get the hearer to ϕ. But to ascribe this confusion to Searle is to require us to think only that in the course of the single sentence quoted in the second paragraph of this paper, Searle passed from believing that giving permission was *not* strictly speaking trying to get someone to do something to the view that, since it was an illocutionary negation (or 'denegation' — [3], p. 4) of a directive, it *was* trying to do so after all. I am reluctant to think this. The trouble, surely, is conflicting classificatory criteria, not straightforward self-contradiction. However, we settle the diagnosis though, there can be no doubt something has gone wrong. Illocutionary point is the official criterion for distinguishing illocutionary act types, and by that criterion permissions are not directives.

II

What about the claims Searle makes in connection with his two subsidiary criteria? Does (3) in fact have the same syntactic structure as standard directives like

(4) I order you to ϕ?

And are permissions in fact illocutionary negations of standard directives like

(5) I order you not to ϕ

and

(6) I forbid you to ϕ?

The answer to the first question is uncertain, if only because it seems that one can retroactively give permission but not retroactively order. However, this is a controversial point which I shall not pursue. It is the second question I wish to consider: are permissions to ϕ in fact negations of orders not to ϕ and of forbiddings to ϕ?

Tsohatzidis claims they are not. I agree, but I do not think his arguments show it. He has two arguments, which I shall label *A* and *B*.

Argument A claims that substituting 'I give you permission' for 'I do not forbid you' or 'I do not order you not' would in some cases result in nonsense. Thus

(7) I do not forbid you to go at all

and

(8) I do not order you not to go at all

would be equivalent in meaning on Searle's view to

(9) *I give you permission to go at all,

which they clearly are not. They cannot be, Tsohatzidis points out, since (9) is ungrammatical while (7) and (8) are not. This observation is, of course, correct, but it does not show Searle is wrong. For Tsohatzidis's objection is based on superficial idiomatic features of

English which are irrelevant to the question whether the equivalences that Searle alleges obtain really do so. There are two ways in which we can understand (7) and (8). The first is by linking the phrase 'at all' to the word 'not', in which case the two sentences are stylistic variants of

(10) I do not at all forbid you to go

and

(11) I do not at all order you not to go.

'Not at all' in this case is just an emphatic negative, and Searle might reasonably reply to Tsohatzidis that (7) and (8) are equivalent not to (9), but to the perfectly grammatical

(12) I by all means give you permission to go,

since 'by all means' seems to be a positive idiomatic analogue of the negative 'not at all'. Hence on this interpretation Tsohatzidis's objection fails.

The second interpretation of (7) and (8) is less natural. It involves linking 'at all' with 'to go'. One may imagine a speaker saying, for instance, 'It is not that I do not forbid you to go on Friday, or to go when you are married. What I do not forbid you to do is to go at all'. That would be a, rather strained, way of saying that it was not just *certain* styles of going that the speaker did not forbid, but all of them. If we take 'at all' in this way, (7) and (8) turn out to be equivalent to, respectively,

(13) I do not forbid you to go, whatever the style of your going

and

(14) I do not order you not to go, whatever the style of your going.

And Searle can reasonably reply to the objection under this interpretation that (7) and (8) are equivalent, not to (9), but to the perfectly grammatical

(15) I give you permission to go, whatever the style of your going.

Hence on this interpretation too, Tsohatzidis's objection fails. Let us consider *Argument B* now.

Argument B claims that there are sentences of the form 'I give you permission to ϕ' which have a different meaning from that of the corresponding sentences of the form 'I do not forbid you to ϕ' or 'I do not order you not to ϕ'. Tsohatzidis's example is

(16) I give you permission either to leave this house or to marry my daughter,

which, he claims, differs in meaning from both

(17) I do not forbid you either to leave this house or to marry my daughter

and

(18) I do not order you either not to leave this house or not to marry my daughter.

The difference, according to Tsohatzidis, is that (16) 'can be (and indeed standardly is) taken to mean that its hearer will be held responsible for not satisfying the speaker's stated wish if he stays at his house *without* marrying his daugher', while (17) and (18) cannot be understood in this way ([4], p. 82).

Now all three of these sentences are remarkably odd. It is difficult to imagine a use for them in everyday language. And that should make us cautious about drawing conclusions from them. But in any case, the argument is confused. In the first place, notice that Tsohatzidis thinks (16) *states a wish* that the hearer should do something. But to state a wish to someone that he should do something, when it is understood that the hearer will be held *responsible* for not fulfilling the wish, is surely to perform a directive illocutionary act; it is to try to get him to do it. Hence giving permission would turn out in this case to be a directive after all — which is inconsistent with Tsohatzidis's view. But in the second place, whatever (16) may *imply* about the speaker's wishes, it certainly does not *state* that he has them; it appears simply to grant the hearer permission to do one, or possibly both, of two actions. And, correspondingly, (17) and (18) are apparently (declarations of) refrainings from forbidding the hearer, or from ordering him not, to do one, or possibly both, of two actions. If we distinguish carefully between what an utterance means and what uttering it implies, then, Tsohatzidis's objection is unsuccessful. Searle can reply that (17) and (18) are indeed equivalent in meaning to (16). The reason why Tsohatzidis's objection may look plausible at first sight is that, given our expectations about the sort of context in which (16) might be uttered, we are inclined to take, not *it*, but the *speaker* to mean

(19 I refuse you permission to stay in this house unless you marry my daugher.

In other words, we think the speaker has not said what he meant to say, that utterance-meaning and utterer's meaning have drifted apart. And this may be why Tsohatzidis was misled into thinking that (16) stated the speaker's wish; he was reading a further assumed utterer's meaning into the utterance itself.

III

Both Tsohatzidis's arguments fail, then. Nevertheless the Searlean analysis is mistaken; (3) is not equivalent to either (1) or (2).

Perhaps, as Tsohatzidis further argues ([4], pp. 82–3), (1) is not equivalent to (2) either; but certainly declaring that one does not order that-not, or that one does not forbid, is not the same as giving permission. We can establish that this is so by showing that the conjunction of (1) or (2) with the negation of (3) is in no way inconsistent. Suppose then that a subordinate proposes a course of action to his superior, the merits of which the superior is unable to determine. At the end of their meeting the superior declares

> (20) I do not order you not (forbid you) to do it, and I do not give you permission to do it. I need more time to consider it.

On the Searlean analysis uttering the first conjunct would be to give permission while uttering the second would be to withhold it. Clearly this is not the case; the superior is merely abstaining from making a decision whether to forbid or to permit, not inconsistently both giving and withholding permission.

The word 'abstaining' suggests a useful analogy here. Giving permission for an action to be done stands to declaring one does not order it not to be done, or does not forbid it, as voting for a proposal stands to merely not voting against it. If I do not vote against a proposal, I expose myself less than if I vote for it, just as if I refrain from ordering an action not to be done, or from forbidding it, I expose myself less than if I give permission for it to be done. If I merely do not vote against the proposal to go hang-gliding, I incur less responsibility for the consequences should hang-gliding turn out to be a disaster than I incur if I vote for it. And if I merely declare I do not order someone not to go shooting, or do not forbid it, I incur less responsibility for the consequences should going shooting prove to be a calamity than I incur if I give him permission to go shooting. Another analogy would be between not denying that a proposition is true and assenting to its truth.

These analogies indicate that the relation of giving permission to not ordering that-not, or not forbidding, on the one hand, and of ordering that-not, or forbidding, to not giving permission on the other, is that in each case someone who performs the former act cannot consistently refuse to perform the latter, while someone who performs the latter can consistently refuse to perform the former. If I give someone permission to ϕ, I am *eo ipso* committed to not ordering him not to ϕ, or not forbidding him; but, as we have seen, if I do not order him not to ϕ, or do not forbid him, I am not *eo ipso* commited to giving him permission to ϕ. And if I order him not to ϕ, or forbid him, I am *eo ipso* committed to not giving him permission to ϕ; but if I do not give him permission to ϕ, I am not *eo ipso* committed to ordering him not to ϕ or to forbidding him.

IV

I now turn briefly to an issue raised by the rejection of the Searlean classification of permissions as directives. If they are not directives, in which of Searle's five categories of illocutionary act do they belong? It seems that the only plausible allocation of them is to either commissives or declaratives. But the Searlean account of these two categories excludes permissions from each of them. On the one hand, commissives are 'by definition about the future actions of the speaker', which lays 'precise constraints on the propositional content: the proposition must represent some future course of action by the speaker' ([3], p. 55). But, according to Searle ([2], p. 22), the propositional content of

(21) I permit you to mow the lawn

would be

You will mow the lawn,

which is not a future action of the speaker. So permissions are not commissives. On the other hand, 'By definition a declarative illocution is successful only if the speaker brings about the state of affairs represented by its propositional content... All successful declaratives have a true propositional content' ([3], p. 57). But, according to Searle, the propositional content of

(22) I permit you to kiss my wife

would be

You will kiss my wife,

which, since you may not avail yourself of my permission, is not made true by the successful utterance of (22). So permissions are not declaratives either. (This is not to deny — or endorse — the later Searlean view ([3], p. 3) that *all* performative sentences have the illocutionary force of a declaration, the propositional content of which is that the speaker performs the illocutionary act named by the performative verb. Whatever we may think of that view, what we are concerned with here is only the propositional content on which the performative verb 'permit' operates.)

Thus it appears that the existing Searlean categories cannot accommodate permissions. This is a serious matter for the Searlean theory, since it makes the strong claim that 'there are five and only five fundamental types' of illocutionary force, and thus five and only five illocutionary ways of using language ([3], p. 52). It seems that this claim can be maintained only if either the account of the propositional content of permissions or the accounts of the different categories of illocutionary acts are revised.[1]

University of Hong Kong
Pokfulam Road, Hong Kong

© Christopher New 1988

REFERENCES

[1] J. L. Austin, *How to do things with Words*, Oxford: Clarendon Press, 1962.
[2] J. R. Searle, *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge: Cambridge University Press, 1979.
[3] J. R. Searle and D. Vanderveken, *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press, 1985.
[4] Savas L. Tsohatzidis, 'Deontic Trouble in Speech Act Botany', *Analysis* 47.2 (March 1987), 80c3.

---

[1] This paper evolved to its present form through adaptation to comments by members of the Philosophy Department Seminar, University of Hong Kong, and by the Editor.

MORALLY ACCEPTABLE DIVESTITURE

By KERRY S. WALTERS

STEVEN M. CAHN (ANALYSIS 47.3, June 1987, pp. 175–6) worries over whether principled divestiture of tainted stock is possible. Suppose I own stock in a corporation that's embarked on a policy I consider immoral. I therefore wish to dissociate myself from the corporation. To free myself of complicity I must sell the stock, which entails that a third party purchase it. But this 'entangles' the third party in a situation I deem reprehensible. And to knowingly entangle a third party in such a situation is likewise reprehensible, Cahn suggests, even if I apprise him of my reasons for divestiture. How then is principled divestiture possible?

The following analogy sheds light on Cahn's puzzle. Suppose I'm employed by a firm that's embarked on a policy I consider immoral. I therefore wish to dissociate myself from the firm. This entails my resignation. Moreover, my employer assures me I will be replaced. But one of the conditions of my contract, which I freely accepted at the time of appointment and still see as morally binding, is as follows. If I give intent to resign, I may legitimately step aside only after the firm has found my replacement and I have cooperated in his on the job training for a stipulated length of time. As a condition of employment, my replacement presumably will be required to acquiesce to or at least not oppose the firm's objectionable policy. Consequently, my resignation 'entangles' him in an immoral situation, and this is reprehensible. How then is principled resignation possible?[1]

Surely no one wants to suggest I'm duty bound not to tender my resignation merely because a third party will take my place. The ideal course of action, of course, might be for me to remain in the firm's employment and as an insider lobby against the immoral policy. That way no third party would become an accessory through my resignation. But such an action is supererogatory. If remaining goes against the grain, I may legitimately leave with no violation of duty if I conscientiously apprise my replacement of my negative evaluation of the firm. If he still willingly hires on, I can hardly be said to have 'entangled' him. My resignation is a necessary condition for his replacing me, but not for his ethical appraisal of and subsequent response to the firm's immoral policy.

Similarly with divestiture. My ideal recourse might be to retain my stock, donate the dividends to charity and actively lobby for the reformation of corporate policy. True, I will remain a reluctant accessory in what I take to be an immoral situation, but at least I will avoid 'entanglement' of a third party. But I am under no obligation to retain the stock. Doing so makes me an incredibly decent fellow. It shows I'm willing to sacrifice my peace of mind for the sake of others. But it is clearly supererogatory. If it too abuses my conscience or endurance, I may sell with no violation of duty, provided I conscientiously apprise the buyer of my reasons for divestiture. If he still willingly buys, either because he disagrees with my ethical reservations or agrees without giving a damn, I can hardly be said to have 'entangled' him. My selling is a necessary condition for his purchase, but not for his ethical response.

Is this analysis merely a smokescreen for ethical buck-passing? I don't think so. I'm not claiming divestiture leaves me with spotless hands. I am, after all, retreating rather than redressing. But neither does it totally condemn me. Granted, divesting is a less than ideal recourse. But in light of the predicament's difficulty, it is morally acceptable. By morally acceptable I mean that course of action which falls within the bounds of ethical propriety without being fully virtuous. Given the complexity of certain ethical quandaries, it may be unlikely if not impossible to achieve a completely satisfying resolution. But this doesn't mean no resolution is possible. A careful examination may reveal courses of action which are acceptable although not what, in the best case scenario, one would desire. In short, shooting for moral acceptability is making the best of a bad situation. It may leave one with slightly soiled hands, but it may also be the best one can do under the circumstances.

---

[1] My colleague Chan Coulter has correctly pointed out that this analogy isn't perfect. If a shareholder wishes to sell, there is a logical entailment that he involve a buyer. If an employee wishes to resign, it is probable but not necessary that a replacement becomes involved. Still, I don't think the analogy fatally limps, since the employee wishing to resign here considers himself under an ethical obligation to wait out his replacement. That is, a morally necessary condition for his leaving is that his position be filled by a third party.

Is, then, principled divestiture possible? If by 'principled' one means fully virtuous and completely satisfying, probably not. Moreover I suspect no — or at least very few — actions are principled in this strong sense. But if one more realistically takes 'principled' to mean morally acceptable, then yes. Principled divestiture is possible.

Gettysburg College
Gettysburg, PA 17325-1486
U.S.A.

© KERRY S. WALTERS 1988

THE CONTINGENT *A PRIORI*: A REPLY

By TIMOTHY WILLIAMSON

IN 'Reference and Contingency', *Monist* 62 (1979) pp. 161–89, Gareth Evans tried to explain how *a priori* knowledge of contingent truths is possible by arguing that indexical features of the truths in question make them contingent only in a superficial sense. In 'The Contingent *A Priori*: Has it Anything to do with Indexicals?', ANALYSIS 46.3, June 1986, pp. 113–7, I suggested that his explanation is inadequate, since *a priori* knowledge is also possible of deeply contingent truths with no relevant indexical features. Graham Oppy disagrees; in 'Williamson and the Contingent *A Priori*', ANALYSIS 47.4, October 1987, pp. 188–93, he claims to have detected hidden indexicality in my examples.

I argued that beliefs formed by the following method (M) constitute *a priori* knowledge:

(M) Given a valid deduction from the premiss that someone believes that *P* to the conclusion that *P*, believe that *P*.

If one puts 'There is at least one believer' for '*P*', the *a priori* knowledge is of a non-indexical contingent truth. It will not be necessary to rehearse the details of this argument in order to explain why I do not find Oppy's objections persuasive.

One problem with Oppy's discussion is that most of it deals not with (M) but with the following rule of inference not mentioned in my paper:

$$\text{APK} \qquad \begin{array}{l} Bxp \rightarrow p \\ \therefore p \end{array}$$

Here '*Bxp*' means '*x* believes that *p*'; since Oppy also uses '→' in formalizing MP, it is presumably the material conditional. He seems to read (M) as involving the use of APK. My claim, it is said, would then be that APK is a 'pragmatically justified rule of inference – i.e. APK is analogous to the rule of MP . . . except that it is only truth-preserving *when used*' (p. 190). Oppy toys with the objection that APK would not be truth-preserving when used by a machine in a subject-less world (ibid.; he does not explain how a computer not programmed by any subject nor interacting with one and by hypothesis incapable of beliefs can use inference rules and input and output truths and falsehoods). Now, of course, APK would not be truth-preserving even when used by an ordinary believer. For one can use inference rules to go from one supposition to another as well as from one belief to another; if *p* is false and believed by no one, APK will take you from the vacuously true supposition $Bxp \rightarrow p$ to the false supposition *p*. The best that can be said for APK is that if *x* has the true belief $Bxp \rightarrow p$ and *x* has the belief *p* then the latter is true too. Thus APK, far from clarifying (M), needs to be regimented into a similar form:

(MAPK) Given the belief that if someone believes that *P* then *P*, believe that *P*.

My argument could have been cast in terms of (MAPK), although it was not. The idea would be that (MAPK) yields *a priori* knowledge as output given *a priori* knowledge as input; thus it would process *a priori* knowledge of the necessary truth that if someone believes that there is at least one believer then there is at least one believer into *a priori* knowledge of the contingent truth that there is at least one believer. Both (M) and (MAPK) should be thought of as something like abstract mechanisms for generating beliefs, not as inference rules. There are certainly empirical questions about the realization of such mechanisms within human psychology, but that is another issue.

Oppy's main objection to me (pp. 192–3) is that APK preserves truth only when its premiss and conclusion are evaluated relative to the world in which it is used; hence the conclusion *p* should be qualified by '(in this world)' – an indexical expression. Thus the argument would yield *a priori* knowledge only of the indexical (and necessary) truth

(1a) There is at least one believer (in this world).

What Oppy does not point out is that on Evans's account, some version of which he is defending, we can readily have *a priori* knowledge of

(1b) If there is at least one believer (in this world) then there is at least one believer.

Since use of *modus ponens* preserves *a priori* knowledge, one could know *a priori* by inference

(1) There is at least one believer.

(1) is my original instance of a non-indexical contingent truth known *a priori*. Of course, this route to knowledge of (1) involves indexicality, but that would scarcely save Evans's explanation; (1) remains deeply contingent. The point is academic, however, since Oppy's indexical qualification of APK anyway fails to make it truth-preserving; the earlier remark about suppositions still applies. APK should be replaced by (MAPK), which has no need of indexical qualifications such as '(in this world)'; to believe that *P* is to believe it about one's own world (which is not to say that all beliefs are indexical). Similarly, (M) stands in no need of indexical qualifications.

If, following Oppy, one tries to make (MAPK) generate the belief that there is at least one believer about a subject-less world *w*, one will presumably put 'In *w* there is at least one believer' for '*P*'; but (MAPK) will then output the false belief that in *w* there is at least one believer only when given as input the false belief that if someone believes that in *w* there is at least one believer then in *w* there is at least one believer; garbage in equals garbage out. One might well be worried about (MAPK) if one confused it with a method for forming beliefs such as:

(MAPK′) Given a belief that in all *F* worlds if someone believes that *P* then *P*, form the belief that in all *F* worlds *P*.

Given as input the vacuously true belief that in all believer-less worlds if someone believes that there is at least one believer then there is at least one believer, (MAPK′) does indeed yield as output the false belief that in all believer-less worlds there is at least one believer. Needless to say, (MAPK) can be truth-preserving even though (MAPK′) is not.

I thus find no reason to modify my original argument. One other question raised by Oppy (at pp. 190–1) deserves to be mentioned. Can I *understand* (1) without already knowing that I am a believer? (There is no need to restrict the question, as Oppy does, to worlds in which I am the only subject.) He does not press the issue, since a negative answer is not obviously correct. What should also be noted is that even if a negative answer were correct, it would constitute no objection to my argument. For it would not show my *a priori* knowledge (with understanding) of the contingent truth that there is at least one believer to be indexical-dependent in any sense not applicable to my *a priori* knowledge of the necessary truth that all believers are believers. My ability to know that there is at least one believer by existential generalization from my knowledge that I exist does not undercut my ability to know it *a priori* by means of (M).

Trinity College,
Dublin 2, Ireland.

© TIMOTHY WILLIAMSON 1988

REPLY TO LOWE ON SHIPS AND STRUCTURES

By HAROLD W. NOONAN

IN a previous reply to E. J. Lowe (ANALYSIS 46.4, October 1986, pp. 218–21) I maintained that the collection of planks constituting a ship at a certain time would itself then qualify as (predicatively) a ship, notwithstanding its non-identity with the ship it constituted, and in support of this contention I cited (following R. M. Chisholm) Webster's definition of a ship as 'a structure used for transportation in water'.

In his 'Reply to Noonan' (ANALYSIS 47.4, October 1987, pp. 201–3) Lowe does not dispute the correctness of this definition, but he denies, appealing to Webster's own definition of 'structure', that it supports my claim. He goes on to say, however, that *whatever* the dictionary said he would dispute my claim, for 'is it or is it not the case that for something to qualify as a ship it should be capable of surviving at least some change in its component parts – such as its planks – through renovation or repair? Clearly the answer must be that it is. *Of course*, a ship may lose or gain a plank without thereby ceasing to be. By contrast, the very identity of a collection of planks is determined by the identity of the planks that are its members . . . and consequently it does not make sense to speak of such a collection surviving the loss or gain of a plank. Hence no such collection will ever qualify as a ship.'

This argument suggests that in order to qualify as a ship at a time (or generally, to satisfy a sortal predicate at a time), an entity must satisfy certain *modal* constraints, i.e. certain modal predicates must be true of it, for example, 'is capable of surviving some change in its component parts'. And doubtless Lowe would also be willing to add that it is equally a necessary condition of an entity's qualifying as a ship at a time that it also satisfy certain historical constraints, i.e. that certain past- or future-reflecting predicates (to use Chisholm's terminology) be true of it. Now if so my claim is easy to resist: for certainly the collection of planks constituting a ship at a time is bound to differ from the ship it constitutes at that time *either* in respect of the modal predicates true of it *or* in respect of the historical predicates true of it (otherwise there would be no reason to deny its identity with the ship), so all that an opponent need do is to point to some predicate (e.g. 'is capable of undergoing change of parts' or 'is *always* a ship'), satisfied by the ship but not by the plank-collection, and claim that its satisfaction is a necessary condition of qualifying as a ship.

But in fact I wish to deny that satisfaction of any such constraint is a necessary condition of being a ship at a time.

Let us look first at modal and then at historical constraints.

Lowe says that it is a necessary condition of an entity's qualifying as a ship that it be capable of undergoing a change of parts. But consider this familiar type of example (taken with irrelevant changes from David Lewis *On the Plurality of Worlds*, Basil Blackwell, 1986, pp. 252–3). A piece of plastic constitutes a toy ship (or boat, for those who care about the difference) throughout its existence, for it is synthesized in the mould in which the boat is formed, and it is destroyed when the boat is destroyed. But there are modal predicates true of the boat, but not true of the piece of plastic and vice-versa. For example, the piece of plastic could have been squeezed into a ball and not destroyed, but this is not true of the boat, and the boat could have undergone repair and replacement of parts, but this is not true of the piece of plastic.

However, if we regard these modal differences as providing us with a reason to deny that the piece of plastic is predicatively a boat we must say that there are *two* entities where the boat is — one a boat-shaped piece of plastic and the other a boat. But to say this would manifest a bad case of double vision — there is manifestly just *one* entity there, which we can refer to indifferently *either* as 'a piece of plastic' *or* as 'a boat'.

It might be replied that given the difference in modal predications true of them we *must* distinguish the piece of plastic and the boat. But this is not so. Instead we can deny that the reference of a modal predicate is independent of the linguistic context in which it occurs, and in particular is independent of the singular term to which it is attached — and then the argument is blocked (for details see Lewis *The Plurality of Worlds*, last chapter, or my 'The Closest Continuer Theory of Identity', *Inquiry* 28 (1985)). However, given that we *can* block the argument in this way, and given the absurdity of denying the identity of the piece of plastic and the toy boat, I submit that we *have* to do it. It cannot then be argued, as Lowe maintains, that an entity cannot qualify as a ship (or boat) unless it is capable of a change of parts. For to argue in this way commits one to an absurd view of the number of entities present in situations of the type Lewis describes.

I have no similar argument to bring against the contention that it is a necessary condition of an entity's being a ship at a time that it have a certain type of history. Of course, many predicates do apply to an entity at a time only if certain things have been or will be true of that entity at other times, for example, 'is a widow' or 'is a future President'. I do not, however, believe that this is true of sortal predicates, or, in particular, true of the predicate 'is a ship', and I think that Webster's definition supports me in this.

Lowe denies that this is so since he takes Webster's definition of a 'structure' to imply that for an entity to be a structure at a time it must not only be structured *at that time*, but must be a structure *at all times* and indeed must be *essentially* a structure. But I do not read the definition as having this implication (it is certainly not, as Lowe suggests, what 'one might expect of a definition' – how then could the dictionary define 'bachelor'?). Anyway I have already argued that it is incorrect to suppose that an entity can satisfy a sortal predicate only if it satisfies such a modal constraint.

But suppose that I am right and a collection of planks constituting a ship at a time may itself then qualify as predicatively a ship in virtue of the non-modal, non-historical properties it then possesses. What unwelcome consequences follow? None, I think. We must then, I have argued, distinguish between restricted and unrestricted quantification over things which are predicatively a ship, and hence we must distinguish between two senses of the predicate 'is a ship', i.e. 'is predicatively a ship' and 'is identical with some ship', only one of which (the former) applies to collections of planks constituting ships. But these distinctions can be made without falling into inconsistency. Nor is someone adopting Lowe's position able to claim that his account is more economical, for he too must distinguish between two senses of the predicate 'is a ship', namely 'is predicatively a ship' and 'is constitutively a ship', only one of which (the latter) applies to collections of planks constituting ships.

The difference between Lowe and myself is simply that he takes the predicate which is true both of ships and plank-collections to be the semantically complex 'is constitutively a ship', whilst I take it to be the semantically simple 'is predicatively a ship', this difference in turn deriving from the fact that I accept, while Lowe denies, that the non-modal, non-historical properties possessed by an entity at a time may warrant the application to it, at that time, of the simple predicate.

I hope that this note has clarified what is at issue between us.

University of Birmingham
P.O. Box 363, Birmingham B15 2TT

© HAROLD W. NOONAN 1988

WHY ONLY THE BEST IS GOOD ENOUGH

By STEPHEN GROVER

LEIBNIZ claimed that if there were no best of all possible worlds, God would not have created any world at all. This can be defended in the following way: suppose there is no best of all possible worlds, but only an infinite series of possible worlds, each marginally better than the last. An omnipotent God could create any member of this series, whereas a 'weak god', of great but not quite unlimited power, could create many, but not all, of the members of the series. Whichever member of the series God randomly chooses to create, it will be possible that the world He chooses is not as good as the world which a weak god would, if he existed, choose to create. This possibility is incompatible with the claim that God is perfectly good, for that claim must entail the claim that there is no other imaginable agent who could ever act in a way that is morally better than the way in which God acts. It may not impugn God's goodness to say that He could have acted better than He in fact did, but it surely impugns His goodness to say that some other imaginable being could have done so. The easiest way in which this possibility can be excluded is to claim that there is no possible world better than the world which God chooses to create. Either this is the best of all possible worlds, or God is not omnipotent, not perfectly good, or does not exist.

All Souls College,
Oxford OX1 4AL

© STEPHEN GROVER 1988

BLACKWELL

Metaphysical Horror

LESZEK KOLAKOWSKI

work of profound, polemical enquiry, *Metaphysical Horror* explores the nature of philosophy, describing its self-delusions as well as its methods and historical advances.

28 pages, **$16.95** (0 631 15959 2)

On Interpretation

A Critical Analysis

ANNETTE BARNES

Professor Barnes' original account of interpretation makes possible rationality and objectivity in criticism of works of art, arguing that truth has a place in interpretive criticism, even though that place is limited.

176 pages, hardback **$45.00** (0 631 15947 9)
paperback **$16.95** (0 631 15963 0)

Mind, Psychoanalysis and Science

Edited by PETER CLARK and CRISPIN WRIGHT

The distinguished contributors assess psychoanalytic theory and therapy in the light of philosophy of mind, cognitive psychology and current scientific understanding of the workings of the brain.

384 pages, **$49.95** (0 631 14544 3)

Aristotle's Ethics

JAMES O. URMSON

'This book will be an excellent introduction to Aristotle's *Ethics*. It is written in a clear and lively style. It brings out the main philosophical issues, and discusses them in an interesting and illuminating way. As an introduction to Aristotle's ideas about ethics it could not be bettered.' *J. L. Ackrill, Brasenose College, Oxford*

146 pages, hardback **$45.00** (0 631 15673 9)
paperback **$12.95** (0 631 15946 0)

Homosexuality

A Philosophical Enquiry

MICHAEL RUSE

Professor Ruse examines homosexuality, both historically and in the present, to provide an informed background to the many political, legal and moral debates which surround the issue.

312 pages, **$19.95** (0 631 15275 X)

Basil Blackwell
108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016.
Toll-free ordering: 1-800-638-3030 (USA)

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary; T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, Jane Heal, R. F. Holland, C. Hookway, C. A. Kirwan, D. H. Mellor, H. W. Noonan, R. Gaita, M. A. Stewart, A. R. White, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscription prices for 1988 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£13.00	£16.00	U.S. $33.00
Individuals	£8.00	£10.00	U.S. $21.00

Single Issue Prices

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£4.30	£5.10	U.S. $10.60
Individuals	£3.10	£3.70	U.S. $6.70

5 APR 1988

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write to the publishers.

CONTRIBUTIONS. Articles submitted for publication should be sent to the Editor, Peter Smith, Department of Philosophy, The University of Sheffield, Sheffield S10 2TN.

Contributors are asked to send two copies of their articles, which should normally not exceed 3,000 words in length. Footnotes should be kept to a minimum, and wherever possible be avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them, and the spacing of symbols within formulae should be carefully attended to. Single quotation marks are to be used except for purposes of internal quotation and "scare" quotes. The *Oxford Dictionary for Writers and Editors* is the preferred authority on questions of spelling, etc.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable